



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

The influence of limiting working memory  
resources on contextual facilitation in  
language processing.

**Oliver W. T. Stewart**  
**B.A. (Hons), MSc.**



THE UNIVERSITY  
*of* EDINBURGH

A thesis submitted in fulfilment of requirements for the degree  
of Doctor of Philosophy

to

Psychology  
School of Philosophy, Psychology and Language Sciences  
University of Edinburgh  
31/12/2013

## Abstract

Language processing is a complex task requiring the integration of many different streams of information. Theorists have considered that working memory plays an important role in language processing and that a reduction in available working memory resources will reduce the efficacy of the system. In debate, however, is whether or not there exists a single pool of resources from which all language processes draw, or if the resource pool is functionally fractionated into modular subsections (e.g. syntactic processing, lexical processing etc.).

This thesis seeks to investigate the role that working memory capacity plays in the utilisation of context to facilitate language processing. We experimentally manipulated the resources available to each participant using a titrated extrinsic memory load (a string of digits the length of which was tailored to each participant). Participants had to maintain the digits in memory while reading target sentences. Using this methodology we conducted six eyetracking experiments to investigate how a reduction of working memory resources influences the use of context in different language processes. Two experiments examined the resolution of syntactic ambiguities (reduced relative clauses); three examined the resolution of lexical ambiguities (balanced homonyms such as *appendix*); and one explored semantic predictability (It was a windy day so the boy went to the park to fly his... *kite*).

General conclusions are hard to draw in the face of variable findings. All three experiment areas (syntactic, lexical, and semantic) show that memory loads interact with context, but there is little consistency as to where and how this occurs. In the syntactic experiments we see hints towards a general degradation in context use (supporting Single Resource Theories) whereas in the Lexical and Semantic experiments we see mixed support leaning in the direction of Multiple Resource Theories. Additionally, while individual experiments suggest that limiting working memory resources reduces the role that context plays in guiding both syntactic and lexical ambiguity resolution, more sophisticated statistical investigation indicates that these findings are not reliable. Taken together, the findings of all the experiments lead us to tentatively conclude that imposing limitations on working memory resources can influence the use of context in *some* language processes, but also that that influence is variant, subtle, and hard to statistically detect.

## **Declaration**

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree or professional qualification. The work reported here has been executed by myself, except where due acknowledgment is made in the text. The thesis complies with all regulations for the degree of PhD at the University of Edinburgh and falls below the requisite word limit specified.

Oliver W. T. Stewart

31/12/2013

## Acknowledgements

Firstly, I wish to thank my supervisors Dr. Patrick Sturt, Dr. Amit Dubey, and Dr. Martin Corley for their guidance, advice, and encouragement throughout. They have been a pleasure to work with and provided a highly enjoyable and motivating environment in which to study. I also acknowledge the ESRC for their initial interest in the study and its subsequent funding, as well as their understanding and flexibility in difficult times.

I am very grateful to my parents who have provided unwavering support throughout my academic career and particularly during my PhD. I would not have made it this far without them and cannot express my appreciation enough for their past and ongoing support.

I am indebted to my brother, Dr. Edward Stewart, for programming advice and assistance.

To all the members of my group, past and present, I am thankful for providing a fun and laughter filled office which was a pleasure to go to each day. I make particular mention of Dr. Suzy Moat and Dr. Manon Jones for providing invaluable advice, guidance, and good humour during the early stages of my PhD, and Dr. Ian Finlayson for friendship, and the shared experience of writing up.

To Graham Harper I am indebted for much emotional support, practical and technical advice, and unending humour during the bad times and the good. He is a true friend who has been there for me, unwaveringly, throughout all my adult life.

To those I call friends and acquaintances: I thank you for all the support, the good times, the laughter, and the days spent out on the hills and rocks. These are times I won't forget, and I thank you for sharing them with me.

To Dr. Mehmet Karatay, and his wife Gemma, two of the kindest and most selfless people I have ever met; I thank you for all the support, both emotional and practical, through one of the most difficult and stressful periods of my life.

Finally, I cannot begin to describe my depth of gratitude to Dr. Samantha Capewell, my partner and best friend, who has shared this experience with me more closely than any other could. She has been unwavering in her support, and provided me with inspiration and determination when the pressures of the task were too great. I could not have done it without her.

# Contents

Abstract	i
Declaration	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	viii
List of Tables	xi
List of Abbreviations	xii

## **Section 1: Introduction and Background..... 14**

Chapter 1: Literature Review .....	15
1.1 Long term memory and the short-term store .....	16
1.2 Working Memory – The Tripartite and Multi-Component Models .....	20
1.2.1 Overview of the multi-component model .....	20
1.2.2 The Phonological Loop .....	21
1.2.2.1 Serial Order recall .....	22
1.2.2.2 The Phonological Similarity effect .....	22
1.2.2.3 The Word Length Effect and Articulatory Suppression .....	23
1.2.3 The Visuospatial Sketchpad .....	23
1.2.4 The Central Executive .....	24
1.2.4.1 Task switching .....	26
1.2.4.2 Selective attention and Division of attention .....	27
1.2.5 The Episodic Buffer .....	30
1.2.6 Summary of the multi-component model .....	31

1.3 A Capacity Constrained Perspective and Domain Specificity.....	32
1.3.1 Domain General / Domain Specific Resources.....	35
1.4 Capacity Constraints and Dual-Task paradigms.....	42
1.4.1 Interruptive vs. Non-Interruptive tasks .....	43
1.4.2 Storage, Processing, or Both .....	43
1.4.3 Semantic Interference in word list recall .....	47
1.4.4 Digit Span .....	50
1.5 Overview of the current study .....	51
Chapter 2: Digit Span Task .....	52
2.1 Introduction .....	52
2.2 Creation of Digit Span Task .....	53
2.3 Digit Span Pre-Test and Titration .....	53
<b>Section 2: Cognitive Load, Context, and Syntactic Ambiguity Resolution.....</b>	<b>55</b>
Chapter 3: NoLoad, Context, and Syntactic Ambiguity (Experiment 1).....	66
3.1 Introduction .....	66
3.2 Methods.....	68
3.3 Results.....	75
3.4 Discussion.....	81
Chapter 4: Load, Context, and Syntactic Ambiguity (Experiment 2) .....	82
4.1 Introduction .....	82
4.2 Methods.....	83
4.3 Results.....	85
4.4 Discussion.....	90
Chapter 5: Combined Analysis of NoLoad and Load Experiments.....	92
5.1 Introduction .....	92
5.2 Methods.....	95

5.3 Results .....	96
5.4 Discussion.....	111
Chapter 6: End of Syntax Section - Summary and Discussion.....	114
<b>Section 3: Cognitive Load, Context, and Lexical Ambiguity Resolution .....</b>	<b>120</b>
Chapter 7: Generation of Lexical Items and Materials (Inc. an Online Experiment) .....	129
7.1 Introduction to Materials.....	129
7.2 Acknowledgement of Sources .....	129
7.3 WebExp Online Experiment .....	129
7.3.1 Methods.....	130
7.3.2 Results.....	133
7.3.3 Discussion.....	134
7.4 Unambiguous Controls (British National Corpus).....	134
7.5 Aim and Style of Final Sentences .....	135
Chapter 8: NoLoad, Context, and Lexical Ambiguity (Experiment 3).....	137
8.1 Introduction .....	137
8.2 Methods.....	138
8.3 Results.....	141
8.4 Discussion.....	151
Chapter 9: Load, Context, and Lexical Ambiguity (Experiment 4) .....	154
9.1 Introduction .....	154
9.2 Methods.....	155
9.3 Results.....	156
9.4 Discussion.....	165
Chapter 10: Combined Analysis of NoLoad and Load Experiments.....	168
10.1 Introduction .....	168
10.2 Methods.....	170



10.3 Results .....	171
10.4 Discussion.....	195
Chapter 11: Load, Context, and Lexical Ambiguity - Within Subjects (Exp 5).....	200
11.1 Introduction .....	200
11.2 Methods .....	202
11.3 Results.....	205
11.4 Discussion.....	216
Chapter 12: End of Lexical Section - Summary and Discussion .....	219
<b>Section 4: Cognitive Load, Context, and Semantic Predictability .....</b>	<b>226</b>
Chapter 13: Load, Context, and Semantic Predictability - Within Subjects (Exp 6).....	235
13.1 Introduction .....	235
13.2 Methods .....	237
13.3 Results .....	241
13.4 Discussion.....	252
<b>Section 5: General Discussion.....</b>	<b>257</b>
Aim of Thesis .....	257
Overview and Implications of Empirical Studies.....	258
Syntactic Ambiguity .....	258
Lexical Ambiguity .....	259
Semantic Predictability .....	260
General Discussion .....	261
Domain General or Domain Specific Resources .....	261
Experimental reduction in resources vs. Low span readers .....	263
Strategy and Task Demands.....	264
Future Directions .....	265
Concluding Remarks.....	265

References .....	267
Appendix 1: Materials for Syntactic Ambiguity Experiments .....	289
Appendix 2: Materials for Lexical Ambiguity Experiments .....	299
Appendix 3: Materials for Semantic Prediction Experiment .....	312

## List of Figures

<i>Figure 3.1: Experiment 1- Proportion of Regressions for Regressions Out on the NP2 Region. Error Bars display Standard Error. Context x Ambiguity Interaction. ....</i>	<i>76</i>
<i>Figure 3.2: Experiment 1 - Proportion of Regressions for Regressions In on the 'by' Region. Error Bars display Standard Error. Context x Ambiguity Interaction. ....</i>	<i>77</i>
<i>Figure 4.1: Experiment 2 - Proportion of Fixations for Probability of Fixation on the Spillover Region. Error Bars display Standard Error. Context x Ambiguity Interaction. ....</i>	<i>86</i>
<i>Figure 5.1: Experiment 2 - Proportion of Regressions for Regressions Out of the NP2 Region for the Load experiment. Error bars display standard error. Context x Ambiguity Interaction. ....</i>	<i>94</i>
<i>Figure 5.2: Experiment 2 - Proportion of Regressions for Regressions In to the 'by' Region for the Load experiment. Error bars display standard error. Context x Ambiguity Interaction. ....</i>	<i>94</i>
<i>Figure 5.3: Combined Analysis of Experiments 1 &amp; 2. Proportion of Regressions for Regressions Out on the NP2 Region. Error Bars display Standard Error. Context x Ambiguity Interaction. ....</i>	<i>97</i>
<i>Figure 5.4: Combined Analysis of Experiments 1 &amp; 2. Proportion of Regressions for Regressions Out on the NP2 Region for the No Load and Load experiments. Error Bars display Standard Error. A non-significant Context x Ambiguity x Load Interaction. ....</i>	<i>99</i>
<i>Figure 5.5: Combined Analysis of Experiments 1 &amp; 2. Proportion of Regressions for Regressions In on the 'by' Region for the No Load and Load experiments. Error Bars display Standard Error. A marginal Context x Ambiguity x Load Interaction. ....</i>	<i>100</i>
<i>Figure 5.6: Combined Analysis of Experiments 1 &amp; 2. Mean inspection time for First Pass on the Verb. Error Bars display Standard Error. Context x Load Interaction. ....</i>	<i>103</i>
<i>Figure 5.7: Combined Analysis of Experiments 1 &amp; 2. Proportion of Fixations for Probability of Fixation on the Verb. Error Bars display Standard Error. Ambiguity x Load Interaction. ....</i>	<i>104</i>
<i>Figure 5.8: Combined Analysis of Experiments 1 &amp; 2. Proportion of Regressions for Regressions Out on the Verb. Error Bars display Standard Error. Ambiguity x Load Interaction. ....</i>	<i>105</i>
<i>Figure 8.1: Experiment 3 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context x Ambiguity Interaction. ....</i>	<i>142</i>
<i>Figure 8.2: Experiment 3 - Proportion of Fixations for Probability of Fixation on the Pre-Critical Region. Error Bars display Standard Error. Context x Ambiguity Interaction. ....</i>	<i>143</i>

<i>Figure 8.3: Experiment 3 - Mean fixation durations for First Fixation in the Spillover Region. Error Bars display Standard Error. Ambiguity x Lexical Frequency Interaction. ....</i>	<i>145</i>
<i>Figure 8.4: Experiment 3 - Mean inspection time for Second Pass in the Spillover Region. Error Bars display Standard Error. Ambiguity x Lexical Frequency Interaction. ....</i>	<i>146</i>
<i>Figure 8.5: Experiment 3 - Mean inspection time for Second Pass in the Wrapup Region. Error Bars display Standard Error. Ambiguity x Lexical Frequency Interaction. ....</i>	<i>147</i>
<i>Figure 9.1: Experiment 4 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context, Ambiguity, and Lexical Frequency Interaction. ....</i>	<i>158</i>
<i>Figure 9.2: Experiment 4 - Mean inspection time for First Pass in the Critical Region. Error Bars display Standard Error. Context by Ambiguity Interaction. ....</i>	<i>159</i>
<i>Figure 9.3: Experiment 4 - Mean inspection time for Total Time in the Pre-Critical Region. Error Bars display Standard Error. Context by Ambiguity Interaction - Marginal (0.076).....</i>	<i>160</i>
<i>Figure 9.4: Experiment 4 - Mean inspection time for First Pass in the Critical Region. Error Bars display Standard Error. Ambiguity by Lexical Frequency Interaction. ....</i>	<i>161</i>
<i>Figure 10.1: Combined Analysis of Experiments 3 &amp; 4 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context, Ambiguity, and Cognitive Load Interaction. ....</i>	<i>174</i>
<i>Figure 10.2: Combined Analysis of Experiments 3 &amp; 4 - Proportion of Fixations for Probability of Fixation on the Pre-Critical Region. Error Bars display Standard Error. Context, Ambiguity, and Cognitive Load Interaction. ....</i>	<i>176</i>
<i>Figure 10.3: Combined Analysis of Experiments 3 &amp; 4 - Mean inspection time for First Pass on the Combined Region: Pre-Critical &amp; Critical. Error Bars display Standard Error. Marginal Context, Ambiguity, &amp; Load Interaction. ....</i>	<i>178</i>
<i>Figure 10.4: Combined Analysis of Experiments 3 &amp; 4 - Mean inspection time for First Pass on the Spillover Region. Error Bars display Standard Error. Context, Ambiguity, &amp; Load Interaction. ....</i>	<i>180</i>
<i>Figure 10.5: Combined Analysis of Experiments 3 &amp; 4 - Mean inspection time for Total Time on the Pre-Critical Region. Error Bars display Standard Error. Context, Ambiguity, &amp; Load Interaction. ....</i>	<i>182</i>
<i>Figure 10.6: Combined Analysis of Experiments 3 &amp; 4 - Proportion of Fixations for Probability of Fixation on the Combined Region: Pre-Critical &amp; Critical. Error Bars display Standard Error. Context and Cognitive Load Interaction. ....</i>	<i>185</i>
<i>Figure 10.7: Combined Analysis of Experiments 3 &amp; 4 - Proportion of Fixations for Probability of Fixation on the Wrapup Region. Error Bars display Standard Error. Context and Cognitive Load Interaction. ....</i>	<i>187</i>
<i>Figure 10.8: Combined Analysis of Experiments 3 &amp; 4 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Lexical Frequency and Cognitive Load Interaction. ....</i>	<i>188</i>

<i>Figure 10.9: Combined Analysis of Experiments 3 &amp; 4 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context, Ambiguity, and Frequency Interaction.</i>	190
<i>Figure 11.1: Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context and Cognitive Load Interactions for Between Subjects and Within Subject designs. Ambiguous words only.</i>	206
<i>Figure 11.2: Mean inspection time for First Pass in the Spillover Region. Error Bars display Standard Error. Context and Cognitive Load Interactions for Between Subjects and Within Subject designs. Ambiguous words only.</i>	208
<i>Figure 11.3: Mean inspection time for Total Time on the Pre-Critical Region. Error Bars display Standard Error. Context and Cognitive Load Interactions for Between Subjects and Within Subject designs. Ambiguous words only.</i>	210
<i>Figure 11.4: Experiment 5 – Mean inspection time for First Pass in the Spillover Region. Error Bars display Standard Error. Context by Repetition Interaction.</i>	212
<i>Figure 13.1: Experiment 6 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Marginal Predictability by Load Interaction.</i>	242
<i>Figure 13.2: Experiment 6 – Mean inspection time for First Pass on the Article+Noun Region. Error Bars display Standard Error. Predictability by Frequency Interaction.</i>	244
<i>Figure 13.3: Experiment 6 – Mean Regression Path Duration for Regression Path for the Critical Region. Error Bars display Standard Error. Load by Frequency Interaction.</i>	247
<i>Figure 13.4: Experiment 6 – Mean Fixation Durations for First Fixation of the Critical Region. Error Bars display Standard Error. Load by Frequency Interaction.</i>	248

## List of Tables

<i>Table 3.1: Experiment 1 - Raw numbers of regressions out of the NP2 region, and raw numbers of regressions in to the 'by' region. By condition.</i>	77
<i>Table 3.2: Experiment 1 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if Context is Unsupportive or Supportive. Effects are also marked if Marginal.</i>	79
<i>Table 3.3: Experiment 1 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if the local Syntax is Ambiguous or Unambiguous. Effects are also marked if Marginal.</i>	80
<i>Table 4.1: Experiment 2 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if Context is Unsupportive or Supportive. Effects are also marked if Marginal.</i>	88

<i>Table 4.2: Experiment 2 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if the local Syntax is Ambiguous or Unambiguous. Effects are also marked if Marginal.</i>	89
<i>Table 5.1: Context Effects which were present / absent between experiments. NL = present in NoLoad experiment, L = present in Load experiment. x = not present.</i>	92
<i>Table 5.2: Ambiguity effects which were present / absent between experiments. NL = present in NoLoad experiment, L = present in Load experiment. x = not present.</i>	93
<i>Table 5.3: Combined Analysis of Experiments 1 &amp; 2 – Interactions which are not discussed in the text.</i>	102
<i>Table 5.4: Combined Analysis of Experiments 1 &amp; 2 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if Context is Unsupportive or Supportive. Effects are also marked if Marginal.</i>	108
<i>Table 5.5: Combined Analysis of Experiments 1 &amp; 2 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if the local Syntax is Ambiguous or Unambiguous. Effects are also marked if Marginal.</i>	109
<i>Table 5.6: Combined Analysis of Experiments 1 &amp; 2 - Main Effects of Load. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if participants are under Load or NoLoad. Effects are also marked if Marginal.</i>	110
<i>Table 7.1: Method for calculating Variance Thresholds of 20%, 30% and 40% scaling by number of senses. In this case we have 3 senses, as in the three senses of the word Spirit (ghost, alcohol, fighting spirit).</i>	132
<i>Table 7.2: An example of the application of variance thresholds to the three sense word Spirit.</i>	132
<i>Table 7.3: Words Selected as most balanced as judged by our testable population.</i>	133
<i>Table 7.4: Ambiguous items and their Length and Frequency Matched Controls.</i>	135
<i>Table 8.1: Experiment 3 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if Context is Neutral or Supportive. Effects are also marked if Marginal.</i>	149
<i>Table 8.2: Experiment 3 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if the Critical Noun is Ambiguous or Unambiguous. Effects are also marked if Marginal.</i>	150
<i>Table 9.1: Experiment 4 - Conditional Proportion of Fixations in % for Probability of Fixation: Critical Region. Context, Ambiguity, &amp; Lexical Frequency Interaction.</i>	157
<i>Table 9.2: Experiment 4 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if Context is Neutral or Supportive. Effects are also marked if Marginal.</i>	163

<i>Table 9.3: Experiment 3 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if the Critical Noun is Ambiguous or Unambiguous. Effects are also marked if Marginal.</i>	164
<i>Table 10.1: Context Effects which were present / absent between experiments. NL = present in NoLoad experiment, L = present in Load experiment. x = not present.</i>	168
<i>Table 10.2: Ambiguity effects which were present / absent between experiments. NL = present in NoLoad experiment, L = present in Load experiment. x = not present.</i>	169
<i>Table 10.3: Combined Analysis of Experiments 3 &amp; 4 - Conditional Proportion of Fixations in % for Probability of Fixation: Critical Region. Context, Ambiguity, &amp; Cognitive Load Interaction.</i>	172
<i>Table 10.4: Combined Analysis of Experiments 3 &amp; 4 - Conditional Proportion of Fixations in % for Probability of Fixation: Pre-Critical Region. Context, Ambiguity, &amp; Cognitive Load Interaction.</i>	175
<i>Table 10.5: Combined Analysis of Experiments 3 &amp; 4 - Conditional Mean fixation durations in ms for First Pass: Combined Region: Pre-Critical &amp; Critical. Marginal Context, Ambiguity, &amp; Load Interaction.</i>	177
<i>Table 10.6: Combined Analysis of Experiments 3 &amp; 4 - Conditional Mean fixation durations in ms for First Pass: Spillover Region. Context, Ambiguity, &amp; Load Interaction.</i>	179
<i>Table 10.7: Combined Analysis of Experiments 3 &amp; 4 - Conditional Mean fixation durations in ms for Total Time: Pre-Critical Region. Context, Ambiguity, &amp; Load Interaction.</i>	181
<i>Table 10.8: Combined Analysis - Interactions involving Frequency which are not reported in the Results section.</i>	184
<i>Table 10.9: Combined Analysis of Experiments 3 &amp; 4 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if Context is Neutral or Supportive. Effects are also marked if Marginal.</i>	192
<i>Table 10.10: Combined Analysis of Experiments 3 &amp; 4 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if the Critical Noun is Ambiguous or Unambiguous. Effects are also marked if Marginal.</i>	193
<i>Table 10.11: Combined Analysis of Experiments 3 &amp; 4 - Main Effects of Load. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if participants are under Load or NoLoad. Effects are also marked if Marginal.</i>	194
<i>Table 11.1: Experiment 5 - Interactions involving Frequency and / or Repetition which are not reported in the Results section.</i>	211
<i>Table 11.2: Experiment 5 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if Context is Neutral or Supportive. Effects are also marked if Marginal.</i>	214

<i>Table 11.3: Experiment 5 - Main Effects of Load. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if participants are under Load or NoLoad. Effects are also marked if Marginal. ....</i>	<i>215</i>
<i>Table 13.1: Experiment 6 - Interactions involving Frequency which are not reported in the Results section. ....</i>	<i>246</i>
<i>Table 13.2: Experiment 6 - Main Effects of Predictability. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if the Noun is Unpredicted or Predicted by the preceding context. Effects are also marked if Marginal. ....</i>	<i>250</i>
<i>Table 13.3: Experiment 6 - Main Effects of Load. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times &amp; Lower Proportions of Fixations if participants are under Load or NoLoad. Effects are also marked if Marginal. ....</i>	<i>251</i>

## List of Abbreviations

WM	Working Memory
SR	Single Resource (Theory)
SLIR	Separate Language Interpretation Resource (Theory)
FP	First Pass
SP	Second Pass
RP	Regression Path
TT	Total Time
FF	First Fixation
Pfix	Probability of Fixation
RO	Regressions Out
RI	Regressions In

## Section 1: Introduction and Background

Working memory has been known to be an integral part of the understanding of language since research in the field began. The structures and mechanisms of memory which underpin how language processing is constrained have been studied in detail using a variety of methodologies and techniques. Many questions, however, remain unanswered. One ongoing debate in the literature, and the focus of this thesis, is whether the language comprehension system is supported by a single resource pool which feeds all processes related to turning a language input into meaning, or multiple resource pools which support different processes.

It is the primary intent of this thesis to investigate the Single or Multiple resources question across a variety of different language challenges (i.e. Syntactic ambiguity resolution, Lexical ambiguity resolution, and Semantic prediction) by manipulating the availability of both cognitive resources and supporting contextual information. There are robust findings in the literature pertaining to the use of contextual information normal language processing, and it is our intention to use these as a baseline against which to compare reading behaviour under conditions of reduced resources. Additionally, we note that much of the previous research into this area has been conducted using self-paced reading or fully offline measures such as reasoning or truth verification tasks. We are concerned by the ecological validity and fidelity of these approaches and seek to extend them by providing more natural, and richer data using more modern methodologies. To that end, in the experiments reported here we use fully online eye-tracking rather than self-paced reading or other partially / fully offline tasks. This will allow us to see the full range of reading behaviours (including regressions and skipping behaviours) that may have previously been missed. It is also apparent that much of the research in this area has utilised pre-defined population groups of high and low working memory capacity. Here we take a different approach and rather than relying on the prior-classification of working memory groups we experimentally reduce our participants' working memory capacity using an extrinsic memory load.

Details of the structure of the thesis can be found at the end of the Literature Review chapter (section 1.5).



## Chapter 1: Literature Review

The interface between working memory and language processing has been the subject of scientific scrutiny for many years, beginning with Miller & Chomsky's observation (some 50 years ago in 1963) that doubly centre embedded sentences such as 'the squirrel the dog the cat saw chase climbed the tree' are grammatically viable but nearly impossible to accurately interpret. This posed the question: why, when the rules of grammar are not violated, are these constructions so hard to understand? The answer appeared to be to do with the amount of information being stored and processed and so, to try to answer this question, researchers turned to the Memory literature.

In section 1.1 of the literature review we will follow the path laid out by early research into the area: We will examine the development of the memory literature as it proceeds from an early understanding of memory as a single unitary system, through the fractionation of memory into long-term and short-term components (Ebbinghaus, 1913; Hebb, 1949), and into the concepts of working memory that are prevalent today. Specifically we will trace the developments that led to the introduction of a capacity constrained short-term store (Atkinson & Shiffrin, 1968) and one of the most influential models of working memory: The Tripartite model (Baddeley & Hitch, 1974).

In section 1.2 we will summarise the tripartite model, with brief mention of some of the evidence which led to the development of the structural and functional components of the model, and subsequently the addition of a fourth component (the episodic buffer) to create Multi-Component model (Baddeley, 2000).

In section 1.3 we will refocus the discussion towards to the resource oriented mechanisms behind working memory and language processing, with specific attention being paid to the debate between single resource pool capacity constrained models (Just & Carpenter, 1992) and the fractionation of the verbal component of the working memory system (Waters & Caplan, 1996; Caplan & Waters, 1999). In this section we will consider the approaches and methodologies that have been utilised throughout the past fifty years of working memory and language processing research. This will necessarily involve an investigation of the Individual Differences approach as it has been the principal methodology in driving working memory research.

In section 1.4 we will discuss the place of Dual Task paradigms in working memory research and how they can be used to experimentally reduce the availability of resources. We identify three main categories of tasks and discuss them in terms of their appropriateness for our intended manipulations.

Finally, section 1.5 will summarise the structure of the thesis and give an overview of the remaining chapters.

## 1.1 Long term memory and the short-term store

Given the time frame over which humans have been pondering their ability to uptake and retain information the concept of a short term memory system is relatively novel. Over 2000 years ago, Aristotle considered memory to be like a storehouse, in which all the experiences a person had were kept. This unitary store of information was the generally accepted theory of memory until 1885 when Hermann Ebbinghaus published his revolutionary work: *On Memory*. In this publication Ebbinghaus reported many memory experiments which he had conducted upon himself. He used nonsense syllables (Consonant Vowel Consonant – CVC) as non-semantically related memory items to investigate the processes of learning and forgetting. His most enduring findings were the learning and forgetting curves which respectively describe the timescale of loss and retention of information committed to memory. Within his work on the learning curve Ebbinghaus also identified the serial position effect (the observation that the order in which information is passed into memory has an influence on how quickly it is forgotten and how readily it is recalled), a phenomenon that has been hugely influential in guiding the development of models of working memory (e.g. Baddeley & Hitch, 1974; Rönnerberg, 2003). It was however, his conclusions drawn from the investigation of forgetting that led him posit a componential approach to memory which broke the ‘storehouse’ down into three subsystems: Short term, Long term, and Sensory.

This was the first time Short term memory had been defined separately but it wasn’t until the late 1940s that the idea really took hold when Hebb (1949) defined the mechanisms by which short term and long term memory functioned: short term memory as consisting of temporary electrical activation, and long term memory being driven by neuronal growth. In the late 50s support grew for the distinction Ebbinghaus and Hebb had argued (see Brown, 1958 and Peterson & Peterson, 1959 for some initial testing of short term memory) but in

the early 60s some researchers (e.g. Melton, 1963; Postman, 1964) argued that the memory system was not componential and could not be broken down into long-term and short-term components. Specifically, Melton (1963) argued that long-term memory and short-term memory were merely points on a continuum of the single unitary memory system. As we will discuss towards the end of this section, this argument has held some sway (see for example Craik & Lockhart, 1972; Nairne, 2002) and it is the issues with the componential model (Baddeley & Hitch, 1974) that this perspective highlights (principally the integration of short-term / working memory information with the long-term store) that have driven the development of the current 'standard model' (the four component model proposed by Baddeley in 2000).

Despite the concerns of unitary system proponents like Melton and Postman, by the late 60s models incorporating a division between long-term and short-term memory systems were becoming prevalent and generally accepted. This was largely due to neurological evidence from patients with brain damage and / or surgical lesions. Barbizet (1963), for example, described a defect known as Korsakoff's syndrome which is caused by tissue damage in the hippocampal region. Patients with Korsakoff's are unable to pass events into the long-term store (and so cannot generate new long term memories) but their proficiency on immediate recall tests is largely unaffected. Similarly, Milner (1966) wrote of surgical bilateral lesions in the hippocampal region causing persistent and potent long-term memory deficits, while leaving short-term recall relatively unaffected, and the efficacy of rehearsal processes intact. These examples of biological dissociations between the two systems posed a difficult problem for theorists attempting to maintain the concept of a single unitary memory system, and strongly supported componential models. See Squire & Zola-Morgan (1991) for a review of the neurobiological findings relevant to this issue between 1957 and 1991.

Atkinson & Shiffrin (1968) provided a three part componential framework in which models of memory could be constructed. They posited a distinction between permanent structural components of memory and fluid control components, using the analogy of a computer and programmer to explain their distinction. Structural components amounted to the hardware of the system and program code that the programmer could not edit (e.g. basic memory stores). Control components consisted of instructions the programmer might give, and existing components of code which the programmer could edit (e.g. search strategies and

rehearsal procedures). Following Ebbinghaus and Hebb, they broke down the memory system into the long-term store, the short-term store, and the sensory register. The sensory register briefly handles input but only sustains that information for a few seconds before it decays. The short-term store takes the input from the sensory register and can sustain it for up to thirty seconds before decay starts to occur. However, the decaying information can be sustained further by a control process called Rehearsal, in which the subject can refresh the information in the short-term store by repeating it to themselves (see Brown, 1958, for discussion of suppressing rehearsal). Atkinson & Shiffrin assert that this rehearsal process can keep information alive in the short-term store indefinitely. The long-term store is a permanent repository of information which is copied over from the short-term store. It is worth noting at this stage that Atkinson & Shiffrin consider that a structural component of the short-term store is that it is space in which operations can be performed on the current residing information. In this way we observe one of the early uses of the term 'Working Memory' to describe the system that Cognitive Psychologists know it as today (the term was originally coined by Miller, Galanter, & Pribram (1960) but has been used in other fields to refer to other things (e.g. Animal Cognition (Honig, 1978; Olton, Becker, & Handelmann, 1980), Artificial Intelligence (Newell & Simon, 1972), see Baddeley, 2000; 2010 for details).

The above summary of Atkinson & Shiffrin's framework, although necessarily brief, contains mention of the most contentious assertions. Firstly, the structural distinction between short-term memory and long-term memory continues to court controversy. Craik & Lockhart (1972) discuss at length the inadequacies of the multi-store model account specifically targeting the evidence relating to capacity, coding, and forgetting. Their perspective is that the memory system is structurally unitary and that the existing evidence can be interpreted by a 'levels of processing' account. When a subject processes an item in memory, a memory trace is developed and the durability of that trace is dependent on the level at which the item was processed. Items processed deeply, or in several ways (e.g. semantic, phonemic, associative, etc.) have more durable traces than those processed shallowly or in only one way. Rehearsal, and the maintenance of items in what Atkinson & Shiffrin would refer to as the short-term store, is seen as Type 1 processing: This is the repetition of processing at the same level, which does not increase the durability memory trace. Type 2 processing, is considered to be the deepening of analysis by utilising different levels of processing, and does increase the durability of the memory trace. In this way Craik & Lockhart are able to incorporate the existing evidence relating to incidental learning (see

Mclaughlin, 1965, and Postman, 1964, for comprehensive reviews) as well as that relating to rehearsal and repetition, without falling foul of the biggest threat to componential models: functional dissociations.

Shallice & Warrington (1970) and Warrington & Shallice (1969) report a patient known as K.F. who displays a functional dissociation between short-term and long-term memory retention. K.F. is capable of creating and retaining long-term memories normally, but has a significant deficit in short-term memory capability amounting to a reduction in digit and letter span to two items or less (see Crannell & Parrish, 1957 and Miller, 1956, for early work with digit and letter spans, and Cowan, 2001 for a comprehensive discussion of capacity limitations). Atkinson & Shiffrin's framework posits that the creation of long term memories requires information in the short-term store to be transferred to the long-term store. If K.F. has an impaired short-term store, either in terms of storage, processing, or transfer abilities, then this should impact negatively upon his ability to generate new long-term memories. K.F.'s ability to do this however, is not impaired. In addition, given the previous observations of Baddeley & Warrington (1970), Milner (1966), and Wickelgren (1968), which showed patients with the opposite deficit (impaired long-term, but intact short-term memory), and the observation that the anatomical regions engaged during short-term and long-term store activity are disparate (Warrington & Shallice, 1969); Shallice & Warrington (1970) propose a double dissociation between these two components of memory. Therefore, contrary to the structures and functionality of models such as Atkinson & Shiffrin's, the long-term store must have a mechanism for taking input which does not involve the short-term store. Thus Shallice & Warrington propose an architecture in which input is passed to both the long-term and short-term stores in parallel.

Above we have briefly discussed the development of models of memory, from Aristotle's view of a storehouse, through Ebbinghaus and Hebb's segregation of memory into long-term, short-term and sensory, via Atkinson & Shiffrin's componential framework, with brief mention of Craik & Lockhart's unitary store with differing levels of processing, and finally to Shallice & Warrington's parallel multi-store model. In doing so we have focused principally on evidence elicited from patients with neurobiological dissociations. In the next section we will focus the discussion onto models born of the research above but focusing, via experimental investigation rather than neurobiological observation, on the development of the short-term store.

## 1.2 Working Memory – The Tripartite and Multi-Component Models

“Working memory has been variously defined as a system of interacting components that maintain newly acquired and reactivated stored information, both verbal and nonverbal, and make it available for further information processing.” (Becker & Morris, 1999, p. 1). This definition neatly summarises the conceptualisation of working memory and, on the surface, may sound similar to the description of Atkinson & Shiffrin’s short-term store discussed above. Indeed, in terms of the overarching conceptions of that version of the short-term store and the current ideas of working memory, there is little difference between them. The significant variations come in the underlying structural and functional details.

In 1974 Baddeley & Hitch proposed a tripartite model of working memory which, with the addition of one crucial feature leading to the creation of the Multi-Component model (Baddeley, 2000), remains one of the most popular and well accepted models to date. In this section of our review we will summarise the multi-component model and briefly examine the evidence that led to its creation and development, with particular focus on the components involved in language processing.

### 1.2.1 Overview of the multi-component model

In 1974 Baddeley & Hitch presented a body of work detailing ten experiments which investigated the role of memory in three domains of human processing: Verbal reasoning, language comprehension, and learning (see Baddeley, 1992, for a concise overview). Their goal was to investigate if these tasks shared a common working memory system, and if they did, how that related to models of the short-term store. Having detailed the results of the ten experiments, Baddeley & Hitch concluded that there seemed to be a common working memory system which was involved in each of the three tasks they tested. Importantly for this thesis, they used digit span (1-3 items, and 6 items) as one of a series of ways to investigate working memory and found that, while it did impact the results significantly, the level of disruption “was far from massive” (Baddeley & Hitch, 1974, p. 75). They concluded that while the mechanisms responsible for digit span overlapped with a portion of working memory, there was a significant portion of working memory which was not involved in the storage and recall of digits. Following this assertion they went on to suggest that working

memory could be considered a limited capacity workspace that can, within certain predefined limitations, be dynamically assigned both storage and processing tasks.

Baddeley & Hitch posited a three component system in which there is a central executive and two semi-automated slave systems: the phonemic buffer (later termed the phonological loop), and the visual system (later termed the visuospatial sketchpad). The central executive is not taxed unduly by normal verbal or visual processing as this is handled by the respective slave systems. The exception to this occurs if the predefined storage buffers in the slave systems are exceeded, in which case the central executive must make up some of the storage and retrieval deficit (potentially using mechanisms such as chunking (Miller, 1956; Slak, 1970, 1971)). This leads to a reduction in general processing capability as part of the workspace is now consumed by storage demands. Subsequently, Baddeley (2000) introduced a fourth component: The Episodic Buffer.

### 1.2.2 The Phonological Loop

The multi-component model, while being a general model of working memory, has been investigated principally with language processing in mind. The phonological loop is the component of the model which has received the most empirical attention, being both experimentally tractable and language focused. The structure of this slave system is a combination of the sensory and short-term store components of the Atkinson & Shiffrin (1968), with the addition of the Rehearsal control process. It consists of a limited capacity phonological store which can maintain memory traces for a short amount of time and allows for these traces to be refreshed via the process of rehearsal (sub-vocal repetition). The capacity limitation is due to an inability to sub-vocally rehearse a list of items as quickly as the memory trace decays, i.e. a subject might not be able to rehearse the ninth digit of a number such as 186493275, before the first has totally decayed. It is worth noting at this stage that the name 'phonological loop' implies an auditory input, but this is not necessarily the case. The loop does, of course, allow auditory input but it also permits input from visual sources, adding the step of Grapheme-Phoneme conversion before parsing the input to the phonological store.

In the subsections below we will briefly touch on the main phenomena that have informed the development of the Phonological Loop as a separate component of the working memory system.

### *1.2.2.1 Serial Order recall*

A question pertinent to any model of working memory is how it handles the recall of items which follow a specific order. In the digit and letter span tasks for example, it is not only important that participants recall correctly the individual digits / letters, but also that they recall them in the order they were presented. While we will be using Digit Span as a memory loading technique, the details of how order information for memory items is stored and retrieved are of little consequence to this thesis so we will not examine them in great depth. We would, however, point the interested reader to literature pertaining to the three main types of model which attempt to describe how the ordering of memory items is achieved. These are chaining models (Ebbinghaus, 1913; Lashley, 1951; Wickelgren, 1965), Ordinal models (Henson, Norris, Page, & Baddeley, 1996; Page & Norris, 1998; Shiffrin & Cook, 1978), and Positional models (Burgess & Hitch, 1992; Conrad, 1965; Henson, 1998; Lee & Estes, 1977).

### *1.2.2.2 The Phonological Similarity effect*

The phonological similarity effect (whereby more errors are made in recalling the order of items held in memory if those items are phonologically similar to each other, e.g. FS (similar), OQ (dissimilar)) is discussed by Conrad & Hull (1964) and Wickelgren (1965) and featured in two experiments testing verbal reasoning in Baddeley & Hitch (1974). The phonological similarity effect was one of the observed phenomena which helped define the structure of early models of working memory but the exact origin of the effect, and why it occurs, is still not universally agreed upon. It has been, and remains, hugely influential in the development of Working Memory models. For example, the phonological similarity effect has been studied to differentiate between models of working memory (Lewandowsky & Farrell, 2008a; Tehan, Hendry, & Kocinski, 2001), to explore retrieval vs. encoding approaches (Copeland & Radvansky, 2001; Macnamara, Moore, & Conway, 2011), and to propose integrated and novel frameworks of serial order recall (Gupta, Lipinski, & Aktunc, 2005).



### *1.2.2.3 The Word Length Effect and Articulatory Suppression*

In 1975, Baddeley, Thomson, and Buchanan identified and tested the Word Length effect (people struggled more to recall a list of long words than a list of short words). They also conducted an experiment in which they removed the subject's ability to conduct rehearsal by having them continuously repeat a specific word or phrase. In this case they had participants counting from one to eight repeatedly, although others have used different methods (e.g. Levy, 1971, had subjects continuously repeating 'Hiya', Vallar & Baddeley, 1982, used 'the' in the same way, and Baddeley, Chincotta, & Adlam, 2001, used several different methods including repetitions of days of the week and months of the year). This process is called articulatory suppression. Baddeley et al. (1975) showed that, while under the influence of articulatory suppression, the word length effect disappeared with both short words and longer words being inhibited equally.

The importance of the word length effect for models of working memory is not to be underestimated: it has been the subject of extensive investigation (see Mueller, Seymour, Kieras, & Meyer, 2003, for a review), and has, and continues to be a hotbed of controversy (Jalbert, Neath, & Surprenant, 2011), but nevertheless it is considered to be a benchmark finding which models of working memory must be able to explain (Lewandowsky & Farrell, 2008b).

In the tripartite model this is explained as an extension of the trace-decay idea: The phonemic buffer is of limited capacity and subjects cannot sequentially rehearse all items on the list before the memory traces of the early items decay. In this instance the increase in the number of syllables per item simply increases the time required to rehearse each item, which subsequently inflates the overall time lapse between rehearsals of any given item on the list. This explanation is controversial (see Service, 1998, and the commentary by Cowan, Nugent, Elliott, & Geer, 2000, for examples) but Baddeley feels that the time based trace-decay explanation is still adequate (Baddeley, 2003).

### *1.2.3 The Visuospatial Sketchpad*

In Baddeley & Hitch (1974) there were no experiments conducted to investigate the visual processing side of working memory. However, the authors made reference to several other

works which informed the issue (e.g. Baddeley, Grant, Wight, & Thomson, 1975; Brooks, 1967, 1968; Kroll, Parks, Parkinson, Bieber, & Johnson, 1970). These studies, building on each other, showed that visual processing was largely unaffected by phonological processing, leading Baddeley & Hitch to conclude that working memory contained a subsystem which handled visual information, and that that subsystem was distinct and separate from the phonological loop. Originally called the visual system, the slave system which handles visual information has been expanded to include spatial information, and is now known as the Visuospatial sketchpad. The distinction between visual and spatial information comes, in part, from Della Sala, Gray, Baddeley, Allamano, & Wilson (1999) who showed a double dissociation between the interference of visual and spatial information on tasks thought to tap each individually. We will not examine this component of the tripartite model in any more depth as it has little relevance to the thesis. Phonological information that is presented visually undergoes grapheme to phoneme conversion (as is mentioned in section 1.2.2) and from that point on is handled by the phonological loop and the central executive.

#### 1.2.4 The Central Executive

The central executive is the least tractable and least studied component of the multi-component model. Baddeley & Hitch (1974) propose it as a flexible limited capacity central processing workspace responsible for loading and retrieving information from slave systems, queuing and initialising rehearsal routines, and operating on information retrieved from buffers. It was considered that no storage automatically takes place in the central executive, but that it could take on a storage role should the phonemic or visual buffer become overloaded. The central executive has been criticised by researchers such as Parkin (1998) for being little more than a homunculus which takes up the slack when the posited slave systems fail to act as expected. Baddeley admits this, referring to it as “a little man who takes all the decisions that are beyond the capacity of the slave systems” (Baddeley, 1998 p. 524), but argues that there is worth in specifying such a system “provided we accept that the homunculus defines the problem area, but is not the solution” (Baddeley, 2003, p. 835).

In an attempt to define the constraints of the central executive Baddeley and colleagues adopted the Norman and Shallice model of supervisory attentional system (SAS) which

attempted to describe the attentional control of action (Norman & Shallice, 1986). According to Norman and Shallice, the majority of human behaviour is controlled by automated processes which are triggered by environmental cues. These processes frequently conflict with each other however, but are easily and automatically resolved by a system called Contention Scheduling. An example of this, cited in Baddeley (2001), is one of overriding the action plan of driving to a particular location, with one of stopping at a red light. The SAS comes into play when conscious attention must be paid to creating and executing an action plan. This occurs in situations when the problems encountered are novel and have not been dealt with before. The SAS is able to combine information from long term memory with environmental cues to create and execute novel action plans to suit new situations.

Using the framework of the SAS, Baddeley (1996) suggested and explored four properties that he felt the central executive must possess: the abilities to 1). Focus attention, 2). Switch attention, 3). Divide attention, and 4). Interface with long term memory. Each of these areas has been studied extensively and Repovs & Baddeley (2006) provide a concise review. With regards to this thesis, the most pertinent of these properties relates to dividing attention. The research impacting this area will be introduced in section 1.2.4.2 and discussed with specific relation to our digit span task. It is worth noting that Baddeley (1996)'s 4<sup>th</sup> requirement (the ability to interface the contents of working memory with that held in long term memory) leads directly to the genesis of the Episodic buffer, the fourth component of the multi-component model which will be discussed in section 1.2.5.

To summarise the current position of the Central Executive: Repovs & Baddeley (2006) believe that it is directly involved in the management of the short term storage components of working memory (i.e. the phonological loop and the visuospatial sketchpad), as well as any processing of the information held in those components. They also credit it with responsibility for directing and managing attentional focus in order to facilitate both processing in complex cognitive tasks, and complex binding or integration of information held in the storage components.

In the subsections below we will briefly touch on some of the work relating to the attentional component of the Central Executive. Namely, task switching, and the division of attention.

#### 1.2.4.1 Task switching

Following Baddeley (1996), much work has been conducted in an attempt to examine one of the assumed central tenets of the central executive: Its role in the control of attentional focus, division, and switching. One of the principal methodologies used in this exploration is the dual-task paradigm. In this paradigm, participants are required to undertake either two tasks simultaneously, or two tasks which are interleaved. This paradigm has been used to investigate many facets of cognition (pertinently, as we have seen, it featured heavily in Baddeley & Hitch's 1974 investigations leading to the generation of the tripartite model). Here we will give a brief overview the research as it impacts the thesis in terms of the experimental design of two of the experiments.

In 1927 Jersild introduced the task-switching paradigm. He contrasted an AAAA BBBB presentation with an ABABABAB presentation, where A represents task 1 and B represents task 2. He conducted a series of experiments using this paradigm and had students time themselves over the duration of each block. Jersild, while performing no statistics, gave one of the first experimental observations of *switch cost*: The observation that that there is a cost (usually measured in reaction time but sometimes in accuracy) associated with switching between two tasks. Performance is better when a single task is the focus of attention (AAAA BBBB) rather than when attention must be switched back and forth (ABABABAB). Spector & Biederman (1976) replicated and extended this finding and concluded that it was a lack of explicit cueing as to the demands of the next task that caused the switch cost to occur. Many variations of the task-switching paradigm have been created to further investigate this phenomenon (e.g. Allport, Styles, & Hsieh, 1994; Meiran, 1996; Rogers & Monsell, 1995; Sudevan & Taylor, 1987), but we direct an interested reader to Monsell (2003) for a brief overview.

The intricacies of task switching are not particularly relevant to this thesis, except in that we wished to avoid any task switching effects in two of our experiments (see Chapter 11 and Chapter 13). To that end we adapted the alternating-runs paradigm of Rogers & Monsell (1995) which involves switching task for a block of iterations, rather than one to one alternation (e.g. AABBAABB rather than ABABABAB). Our adaption was to include a filler item at the start of each block which was designed to both absorb the switch cost and act

as an explicit cue for the task for the upcoming block. This will be discussed further in the relevant experimental chapters.

#### *1.2.4.2 Selective attention and Division of attention*

More integral to this thesis is the division of attention: One of our primary experimental manipulations was to have participants undertake two tasks concurrently, requiring them to divide their attention between the two. According to Nebel and colleagues, the division of attention can be defined as: “The skill to distribute the limited mental resources to different sources of information” (Nebel et al., 2005, p. 760). Inherent in that definition is the idea that mental resources are limited, an idea strongly advanced by Kahneman (1973) (cited in Norman, 1976) and by Norman & Bobrow (1975). We shall return to this assumption both within this section, and later on when discussing capacity constrained explanations of the central executive component of the multi-component model.

Work characterising the attentional system began over a century ago, with William James providing one of the first overviews of the phenomena of controlled attention (James, 1890). More recent work, building from James’ observations, has focused on defining the problem of selective attention and generating models to explain it. The most famous description of the problem comes from Colin Cherry in 1953. Cherry (1953) used the analogy of a cocktail party to illustrate selective attention. How do we attend to the conversation we find ourselves in, while at the same time filtering out noise from other conversations? The analogy has been expanded by others (e.g. Sternberg, 2003) to include factors such as being able keep a semblance of attention focused on the principal conversation while at the same time passively attending other conversations, and being able to perceive instances in which our name is used, even if it occurs in a conversation we are filtering out.

Several models have been proposed to explain the mechanisms behind selective attention. They fall into three main camps: Filter models in which inputs are filtered at a sensory level (Broadbent, 1958; Deutsch & Deutsch, 1963; Moray, 1959; Norman, 1968; Treisman, 1960, 1964), Two-Stage models in which all sensory information is processed initially for physical information (pitch, volume, etc.), and subsequently semantically for meaning if sufficient working memory resources are available (McCann & Johnston, 1992; Neisser, 1967), and

Limited Resource models which state, in a fairly underspecified way, that there is a single pool of attentional resources that can be allocated between all concurrently running attentional tasks (Kahneman, 1973; Navon & Gopher, 1979).

These models attempt to describe the mechanisms behind selective attention (defined as the ability to attend to a specific input while ignoring, partially or completely, different input streams). We now shift our focus onto the literature pertaining to the division of attention (defined as the ability to attend two input streams simultaneously).

One of the most influential works in the early years of divided attention research was conducted by Spelke, Hirst, & Neisser (1976) (but see also Neisser & Becklen, 1975). The authors ran a series of experiments in which two university students were engaged for five one hour sessions a week for 17 weeks. They were required to read passages for comprehension while simultaneously writing words as they were dictated. In a variety of phases Spelke et al. tested the participants' ability to conduct both tasks simultaneously, using as measures: reading time, reading comprehension, dictation rate, and post-test recognition for dictated words. Initially they were simply required to read at a steady pace for comprehension, while at the same time writing down the words dictated. Subsequently, the dictation lists contained words with meaningful relationships (such as rhymes, words from the same semantic category, or words which would make a sentence) but critically, the participants were not told of these relationships. In later sessions participants were made aware that the lists contained words with relationships, and were required to categorise the dictated words by semantic categories, or interrupt the dictator when they noticed a relationship. Spelke et al. reported that throughout the study, the general finding was that, as time went on, participants became better at maintaining a high reading speed and comprehension level, and could even learn to identify semantic and phonological relationships within the dictated words. With each new task introduction the performance on the reading task would reduce, but with practice it would return to normal levels. The authors interpreted this as evidence to support resource allocation models of attentional control, inferring that each task we are confronted with can, via the medium of repeated practice, be gradually automated so that it consumes less cognitive resource. In this way complex tasks can be performed simultaneously and true division of attention can be achieved.

Posner (1978) was one of the first to report quantitative differences for focused and divided attention. He found decrements in reaction times and accuracy when participants were required to identify multiple stimuli at once (cited in Perry & Hodges, 1999). As Perry & Hodges go on to say, a more common paradigm for investigating the division of attention is to have participants undertake two tasks separately, and then concurrently. By comparing performance on each individual task to that of each task when performed together, we can observe the dual-task-decrement: Participants perform less well (in terms of reaction times and accuracy) when they are required to divide their attention between two tasks.

Kahneman's (1973) model of attentional resource allocation provides one explanation of why the dual-task-decrement might occur: simply that there are limited resources available and the resource demands of both tasks exceed that limit. Resources are therefore shared between the two tasks which perform sub-optimally (but see Pashler, 1994a). However, another explanation is that there are processing stage bottlenecks (Single-channel models: McCann & Johnston, 1992; Welford, 1952) or task specific interferences (Multi-processor models: Allport, 1979; McLeod, 1977) which are not related to limited resources. In the Single-channel view each task requires a given stage of the processing 'machinery' to be focused solely on its needs. Any requests for processing from a second task will be postponed until the 'machinery' is finished and available. This can cause processing bottlenecks if two tasks require the same 'machinery' at the same time. This view originated from the observed psychological refractory period (PRP) in which the closer in time two stimuli are presented, the greater the time taken to respond to the second (this is a very brief and specific description of the PRP, for more detailed discussions see Pashler, 1994b; Sigman & Dehaene, 2005; Telford, 1931). In the Multi-processor view the bottleneck is not caused by queuing for particular processing components as there are multiple processors. Instead task similarity and output is more critical: Tasks which require, for example, similar and simultaneous response outputs, will be bottlenecked at the response stage, and will resolve sequentially. Both the bottleneck and resource sharing models have their support from the literature in terms of explaining the dual-task-decrement. It would appear, however, that the general trend is to consider a two-stage approach: Bottlenecks caused by specific task demands account for much of the dual-task-decrement (as is demonstrated by a reduction in the decrement when the modality of the response is

separated (one task requiring a verbal, and one a manual response, for example)), however, once the influence of task specifics is isolated and removed, some decrement remains and must be explained by a limitation of resources (Bourke, Duncan, & Nimmo-Smith, 1996; Pashler, 1990; Tombu & Jolicoeur, 2003).

Much work has been conducted in an attempt to identify which types of tasks interfere with each other, and which do not (see Bourke et al., 1996 for a summary), but other work has begun to examine other factors as well. With a perspective generated from Baddeley's tripartite model, Hegarty, Shah, & Miyake (2000) have investigated the role of the central executive in the division of attention in dual-task situations (see also Ruthruff, Pashler, & Klaassen, 2001). In a series of experiments Hegarty et al. attempted to suppress resources available to the primary task by using a variety of secondary tasks which were designed to stress the central executive to differing levels. Contrary to their expectations, Hegarty et al. found that the more they stressed the central executive, the better the performance on the primary task. The authors explained their unexpected results as being due to a combination of two principal factors. Firstly, the response bottleneck discussed above and secondly a strategic choice on the part of the participants to prioritise one task over the other. In the work reported in this thesis our participants were required to hold information in memory while reading sentences. This divided their attention but the procedure of the experiments did not create a circumstance in which a response selection bottleneck could occur. There was however, a danger of task prioritisation and instructions were given clearly, both explicitly and implicitly through feedback during pretesting, to ensure that each task was considered to be equally as important.

### 1.2.5 The Episodic Buffer

Since the generation of the tripartite model, several shortcomings have been highlighted. Among the more critical are poorly understood interactions between the slave systems and a lack of workspace capable of holding multi-modal information and transferring it between working memory and long term memory. In response to these shortcomings Baddeley proposed a new component called the Episodic buffer (Baddeley, 2000). Drawing from Tulving's definition of episodic memory (Tulving, 1989) (which was in turn drawn from William James' definition of subjective memory some hundred years previous), Baddeley



suggested a limited capacity store which collates information from both of the slave systems and the long term memory store, in one multi-modal workspace. It is controlled by an attentional system (the central executive) which, now that the lack of storage concern has been addressed, can perform binding and integrative functions. Information can be extracted from the buffer by conscious awareness, and the contents of the buffer can, in turn, be guided by conscious intervention. In this way people can bring information from several different sources (including long term memory) into one workspace, and with the influence of conscious control, can manipulate it to perform problem solving and other complex cognitive activities. This also facilitates the transfer of information from working memory into episodic long term memory, which is the basis for experiential learning (see Rudner & Rönnerberg, 2008 for discussion). The episodic buffer can be viewed as either a new component or a fractionation of the abilities and responsibilities that were originally attributed to the central executive.

#### 1.2.6 Summary of the multi-component model

To briefly summarise the multi-component model: Baddeley and colleagues propose a four component model of working memory which comprises an attentional control system called the central executive, two slave systems known as the phonological loop and the visuospatial sketchpad, and a multi-modal workspace called the episodic buffer in which information can be brought together and operated upon.

The Central Executive is an attentional control system which is at least partially consciously controlled. It manages the allocation of attention between the slave systems, and performs binding, problem solving, and other complex cognitive operations upon information held in the Episodic Buffer.

The Episodic Buffer is a limited capacity workspace in which information from the slave systems and long term memory are brought together. It has no active role in performing operations, but instead provides a temporary store in which operations (guided by the central executive) can take place.

The slave systems provide limited capacity, temporary stores of information from sensory inputs. The Phonological Loop handles verbal and auditory inputs, and allows for some subroutines such as rehearsal to facilitate the maintenance of auditory information. The

Visuospatial Sketchpad performs a similar storage function for visual and spatial information.

### 1.3 A Capacity Constrained Perspective and Domain Specificity

Just & Carpenter (1992) present a model which attempts to explain working memory without subdividing the system into componential parts. To place it within the context of this review so far however, the theory which Just and Carpenter present sits most neatly into the component which Baddeley and colleagues would call the Central Executive. The model builds from the capacity constrained model of attentional control posited by Kahneman in 1973, although rather than attention being the resource used to fuel working memory operations, Just and Carpenter specify 'activation'. In this view, both storage tasks and computations utilise the same pool of resources (rather than in the Baddeley model where storage is handled by the slave systems and the episodic buffer, and computations are conducted by the central executive). Each representational element has a level of activation and, as long as that level of activation exceeds a minimum threshold, it is considered to be 'active' in working memory. Active is defined as being available for inspection and computation. If a representational element's level of activation drops below the minimum threshold, it will cease to be available to the working memory system until some new stimuli re-elevates the activation level. There is a finite amount of activation available to the overall system so, as some elements gain activation, others must consequently lose it. This leads to what Just and Carpenter refer to as "forgetting by displacement" (1992, p. 123). Gaining activation occurs through a process of associative rules which propagate from one element to another. Just and Carpenter urge us to consider the example of syntactic processing: We receive a grammatical subject (which becomes active) and, due to the established rules of syntax we predict that a verb will arrive shortly, and so we begin to commit activation to a syntactic verb element. Language processing is assumed to occur in parallel and so several interpretations of the incoming speech stream can be activated and processed simultaneously. This parallel propagation leads to lots of activation being spread across many elements and could easily, if unchecked, exceed the maximum activation threshold of the working memory system. Therefore, if the maximum threshold looks to be exceeded by excessive propagation, the amount that each element can propagate is scaled back. This keeps the levels of activation from exceeding the

maximum threshold, but slows down the processes involved in the current working memory task. Just and Carpenter propose a similar 'scaling back' system for the trade off between storage and computation. If the propagation of the computation being performed on a stored element would exceed the maximum threshold, then the amount of activation for both the computation and storage for that element are scaled back. This has the effect of weakening the storage maintenance for the element, and lengthening the amount of cycles required to produce a propagation. In other words: Should the system become overtaxed processing slows down and storage becomes more fallible.

When processing passages of text, readers will encounter many more elements than can be held in storage. Therefore, suppression mechanisms are in place to reduce the levels of activation of elements which are no longer needed, or no longer relevant. Relevancy takes several forms: Just and Carpenter specifically mention maintaining activation of the most recent and salient clauses, and the most contextually relevant world knowledge (while suppressing older sentences or other knowledge that might have received activation upon initial reading). Suppression can also occur to the results of lower levels of processing once higher levels have been completed (e.g. a syntactic structure once sentential meaning has been extracted). Just and Carpenter view prediction (of upcoming syntactic or semantic elements) as a mechanism in place to reduce the processing demands placed upon the system, and so free up more activation for maintaining stored elements. In their view, related elements gain activation through propagation, which means that less activation is required upon stimuli onset to bring the element into working memory. This forward propagation is particularly relevant to the thesis as it forms a mechanism for integrating context in the process of disambiguating syntactic and lexical ambiguities (see Section 2 and Section 3). It also, at its most extreme, creates a framework in which semantic prediction can occur (see Section 4).

Unlike the multi-component model, in the capacity constraint theory there is no modularity in which certain storage or processing tasks are architecturally separate from others. The observed evidence supporting distinctions is explained as a combination of exceeding capacity limitations, and individual differences varying the amount of capacity available to each individual. From an individual differences perspective, Just and Carpenter suggest two ways in which participants might differ in terms of their working memory capacity. Firstly, some participants might simply have more activation available and so can hold more

elements active, or activate elements more quickly than other people. Secondly, (and less likely in their view) some participants might have more efficient processes (potentially through practice) so each process requires less activation to complete. With the concept of individual differences in working memory span in mind, but not specifically trying to address whether the cause is greater capacity or more efficient processes, Just and Carpenter replicated a famous Ferreira and Clifton study (1986) which showed that participants were unable to utilise contextual information to help resolve syntactic ambiguities as they encountered them. In their replication, to separate the participants into Low Span and High Span groups Just and Carpenter used the Reading Span measure (Daneman & Carpenter, 1980) in which participants are required to read multiple sentences and then recall the final word of each sentence. On average, for the type of sentence Daneman and Carpenter used, participants can correctly recall the final words of between 2 and 5.5 sentences. Just and Carpenter grouped those with a span of four and above into their High Span group, and those with a span of two or below into their Low Span group. The rationale behind the Reading span task is that high span readers will use less capacity processing the sentences and so will have more left over for storage of the final words. Low span readers will exhaust their capacity reading the sentences and will therefore have fewer resources available with which to maintain the final words in memory.

The low span group performed as Ferreira and Clifton had demonstrated, while the high span group appeared able to make use of disambiguating cues early on. The cue to resolving the ambiguity was the same for all participants, leading to the conclusion that the reading span of the participant reflected the amount of capacity available to their working memory system, and the inability to make use of disambiguating cues was due to a lack of available resources rather than any architectural differences between the groups. The perceived modularity of the syntactic processing system (as proposed by Fodor in 1983) is, in this view, merely an artefact of limited resources:

“Capacity constraints have the potential of creating boundaries between different types of processes when the total processing resources are insufficient to permit direct interaction between different processes. Interaction between processes, like other forms of computation, requires operational resources, such as storage of partial products and communication from one process to another. In the absence of resources sufficient to support interaction, two processes that have the requisite interconnectivity may fail to

interact; that is, they may fail to influence each other's ongoing computations. But the boundaries created by capacity constraints are outcomes of resource limitations, and not of architectural barriers." (Just & Carpenter, 1992, p. 125)

In the conclusion of their 1992 paper, Just and Carpenter raise a point which is of particular interest to this thesis. The authors anticipated the question: Are there multiple resource pools upon which language comprehension draws, or is it the case that one single pool of resources feeds all? Just and Carpenter maintain that the data they have discussed are more easily explained in terms of single resource pool available to all language comprehension processes, rather than sub pools available only to specific processes within the language comprehension system. They do state, however, that language comprehension is only one of a great many cognitive processes, and that there is evidence that other groups of processes draw on different pools; language production, arithmetic, and spatial tasks are the given examples.

In the next section we will trace the conflict in the literature regarding the fractionation of the language comprehension resource pool. As has been mentioned, the thesis is concerned with the manipulation of context and the impact of the presence / absence of said context on the resolution of various types of ambiguity, while participants and under a memory load (resources reduced) or not (resources normal). Domain specificity of comprehension resources is important to the interpretation of our results as, if with our memory load manipulation we are depleting a set of resources which are unrelated to those used in contextual integration, or in semantic guidance of ambiguity resolution, we will see a very different pattern of results to if there exists only a single comprehension resource pool.

### 1.3.1 Domain General / Domain Specific Resources

The capacity constrained theory that Just and Carpenter presented caused a conflict in the literature. In 1996, Waters & Caplan published a paper in *Psychological Review* which was aptly titled: *The Capacity Theory of Sentence Comprehension: Critique of Just and Carpenter (1992)*. Waters and Caplan begin by outlining the basics of the Just and Carpenter model before citing neuropsychological data which casts doubt on the claim that the Reading Span task appropriately measures sentence processing resources. The authors then systematically review the evidence provided by Just and Carpenter in support of the

capacity constrained theory and highlight many inconsistencies and problems. Space constraints prevent us from attempting to detail their concerns in any depth, but to summarise they conclude that the capacity constrained model “rests on an unconvincing empirical foundation.” (Waters & Caplan, 1996, p. 769). The authors’ primary concerns about Just and Carpenter’s work are focused in three main areas: Firstly the assertion that the Reading Span task is an appropriate measure of sentence processing resources, secondly methodological issues and interpretations of data concerning syntactic modularity / ambiguity resolution, and thirdly, what they see as a lack of specification in the model leading to limitations in its explanatory power. Waters and Caplan subsequently present an adaptation of the capacity constrained model which alters the central tenet of a single resource pool for sentence processing. The authors suggest fractionating the single resource pool into two component parts: one involved with the processing of input streams into meaningful representations (e.g. syntactic parsing, lexical access, assigning of semantic roles, etc.), and the other available for consciously controlled processing such as lexical search and explicit reasoning. The argument is that individuals with differing working memory spans will perform equally well at sentence comprehension because the task is mostly automated and so falls into the sentence processing resource domain, whereas the measures of working memory span tap the consciously controlled resource pool; there is, therefore, no competition for resources.

Although not part of the peer-review process for the Waters and Caplan paper prior to publication, Just and Carpenter (along with Timothy Keller) wrote a response rejecting many of the criticisms directed towards their model. It was published alongside the critique outlined above. Just, Carpenter, & Keller (1996) rebut many of the criticisms laid against their interpretations of the findings of previous studies, and reassert that the Reading Span task appropriately measures working memory span (and that working memory span appropriately predicts sentence comprehension ability). They also raise some theoretical concerns about the multiple resource pool theory and the validity of the partition between conscious and automatic processing. Their concerns fall into three categories: firstly, evidence showing that factors external to automated sentence processing systems impact sentence processing, secondly the definition of a process as automatic or consciously controlled can change depending on circumstance, and thirdly, automaticity has been shown to be a continuum (e.g. Spelke et al., 1976) whereas the multiple resource pool hypothesis treats it as a dichotomy and has no way to accommodate the gradual

transference from conscious to automatic. Again, we will not delve into the details of each point of contention but as a summary of the intent of the paper: the opening paragraph states that the goal of the paper is to refute “some of Waters and Caplan’s (1996a) incorrect descriptions concerning the empirical support for capacity theory, as well as pointing out the theoretical and empirical difficulties with Waters and Caplan’s alternative hypothesis.” (Just et al., 1996, p. 773). The authors also introduce a new fMRI study which shows additive activation during the reading while performing the Reading Span task, as compared to reading only. Just et al. conclude that this finding (implicating the same brain regions in sentence processing and the Reading Span task) provides strong evidence that the Reading Span task, contrary to Waters and Caplan’s assertions, is an appropriate measure of working memory span in order to test / predict language comprehension ability. See also Just, Carpenter, Keller, Eddy, & Thulborn, (1996) for support that increased neural activation indicates a greater level of resource consumption, which is being engendered by the heightened computational demands of the task.

The question of whether there is a single resource pool from which all sentence comprehension processing is drawn, or whether there are multiple resource pools which different processes draw on separately, was not resolved in this exchange. In 1999, Caplan & Waters published an extremely influential paper titled: Verbal working memory and sentence comprehension. In this paper Caplan & Waters further their proposition that there is a division in the working memory system centred around conscious control. After a comprehensive review of the evidence pertaining to capacity constrained working memory systems, the authors propose a consolidated and refined version of their previous supposition. In their view there are two types of cognitive resources available to the working memory system: Those used for interpretive processing and those used for post-interpretive processing. Interpretive processing consists of any process which assists in transforming “the acoustic signal in a discourse coherent semantic representation” (Caplan & Waters, 1999, p. 93). Among others, interpretive processes include syntactic processing, lexical access and the resolution of causal inferences. These interpretive processes share the common features of being automatic and obligatory. Post-interpretive processes, on the other hand, are consciously controlled and focus more on making use of the semantic representation offered by the interpretive processes. This includes things like remembering a shopping list or resolving a garden path through syntactic reanalysis: Any process which requires slow, conscious manipulation of verbal information. Caplan and Waters refer to

this division as the SLIR (Separate Language Interpretation Resource) hypothesis in contrast to what they are terming Just and Carpenter's capacity theory: The SR (Single Resource) theory. The authors argue that extrinsic memory loads tap only part of the resource pool: they should not interact with language processing at interpretive levels, but would interact with post-interpretive processes. Conversely, the SR theory would predict that an extrinsic memory load would interact with both interpretive and post-interpretive processes as no distinction is made between the two.

Importantly, in this paper Caplan and Waters address two of the main criticisms which have been levied against previous versions of the SLIR. The first criticism they address is the question of what happens when fast, automated processing runs into trouble (e.g. a garden path sentence or one of sufficient complexity to require conscious focus) and interpretive processing becomes slow and conscious (hallmarks of post-interpretive processing). Caplan and Waters suggest that when this happens people simply switch to using the more general post-interpretive resource pool, and point out that there exist models attempting to explain how people might switch from one type of processing to the other (e.g. Sturt & Crocker, 1996 in the realm of syntactic reanalysis). However, recall the work of Spelke and colleagues showing that practice at a task can shift the level of attention required to perform it from fully consciously controlled to something approaching automated. As we have seen above, several researchers have taken this finding to imply that automaticity / consciously controlled processing are points on a continuum, rather than two discreet states. This latter concern is not directly addressed in the paper.

The second issue that is addressed, however, is the idea of modularity / encapsulation. The criticism is that higher level factors have been shown to influence the initial interpretation of a sentence (MacDonald, Pearlmutter, & Seidenberg, 1994; Trueswell, Tanenhaus, & Garnsey, 1994). The argument is that higher level factors should not be able impact the automatic interpretive process, but if they can then it greatly weakens the partition between interpretive and post-interpretive (which, as we have seen, is already in debate). Caplan and Waters do not feel that the influence of higher level factors reduces the distinction between interpretive and post-interpretive processing, arguing that just because a process can take a higher level factor as an input does not mean that the process itself is any less automatic. They argue that "the interpretation process does not have to be



encapsulated to be cognitively separable or to rely on a specialized resource pool.” Thus the separate interpretive resource pool can be maintained.

To wrap up this section we discuss the third perspective on the SR / SLIR debate, which is presented by MacDonald & Christiansen in 2002. MacDonald and Christiansen are quick to point out that they see worth in both the SR and SLIR theories, but that they disagree with the interpretation of the data which led both Just and Carpenter and Waters and Caplan to their respective conclusions. The authors are keen to present a third alternative in which linguistic experience and biological factors are key. MacDonald and Christiansen propose a fully connectionist approach that, by the nature of interactivity, inherently eschews modularity. In this way their approach has more in common with the SR approach of Just and Carpenter than Waters and Caplan’s SLIR approach. However, recall that Just and Carpenter explain individual differences as variation in the overall capacity of participants, rather than the efficiency by which they utilise the capacity they have (Just & Carpenter, 1992, p. 124). In this way the two approaches critically diverge. MacDonald and Christiansen make no distinction between long term language knowledge and language processing, and so they see working memory less as an architecturally defined subcomponent of memory, and more as part of one comprehensive network. Individual differences occur due to varying levels of skill at language processing, developed by experience of verbal and written language processing.

“To the extent that it is useful to talk about working memory within these systems, it is the network itself; it is not some separate entity that can vary independently of the architecture and experience that governs the network’s processing efficiency.” (MacDonald & Christiansen, 2002, p. 38)

MacDonald and Christiansen run simulations through a feed-forward connectionist network in an attempt to show that a). their experiential model can account for the types of findings that Just and Carpenter and Waters and Caplan have debated, and b). that their account is an empirically separable proposition with differing predictions. They also compare their theory to several other accounts which we have not discussed in this review. We will not discuss those comparisons here but mention should be given to Ericsson & Kintsch (1995) who propose a model which straddles the line between the Just and Carpenter account and the MacDonald and Christiansen perspective (being more oriented around the long term language knowledge / experience concept than Just and Carpenter, but less fully interactive

than MacDonald and Christiansen in that long term language knowledge and processing ability are two separable things).

In rebuttal to MacDonald and Christiansen's critique of Capacity Theory, Just & Varma (2002) aim to point out the strong similarities between Capacity Theory and the Experience based connectionist approach. Most of the arguments put forward by Just and Varma centre around refuting the claim that the experiential account is quantifiably different from capacity theory: "substantial portions of the MacDonald and Christiansen article argued about a non-existent difference" (Just & Varma, 2002, p. 56), "in the large, the Macdonald and Christiansen proposal is not different in any important way" (p. 58). Importantly, Just and Varma dispute the criticism that the capacity theory does not include reference to the role of experience in the efficiency of processing. They argue that, despite choosing to couch their theory of individual differences in terms of overall capacity constraints, it could equally as easily be explained in terms of efficiency of language processing skills; something mentioned in the original Just and Carpenter paper (and observed in section 1.3 of this review). Just and Varma also argue that capacity theory does not, despite MacDonald and Christiansen's claims, propose a dichotic separation between knowledge and processing. Instead, they claim that resources and knowledge are inseparable in the sense that the system is non-functional if either is missing. However, in a postscript appended to their 2002 paper, MacDonald and Christiansen respond to Just and Varma's comments, arguing that a). the two models are empirically distinguishable (it is just that the parameters of the capacity constrained model are not well defined enough to bear out the differing predictions), and b). that knowledge and processing are, in their account, not only intertwined but in fact the same thing, whereas in capacity theory they are separable architecturally, if not functionally.

Caplan and Waters (2002) also responded to MacDonald and Christiansen's critique, citing three main areas in which they disagree with the latter's claims. Firstly, that connectionist and production models are significantly structurally different, in terms of specifying a distinction between knowledge and processing. Caplan and Waters argue that even in a fully connectionist model, the structure of the network, which is unaffected by its experiential learning, still constrains its function. Therefore connectionist models do entail some pre-defined architectural constraint, just as production models do, and this implies that the division between knowledge and processing which MacDonald and Christiansen

claim is not present in connectionist models, does still exist. Secondly, that, although the two are highly correlated, the most effective explanation of individual differences in syntactic processing is language experience rather than working memory capacity constraints. Caplan and Waters rebut this claim; they surveyed 40 participants in a working memory study they conducted (20 high and 20 low span individuals), and found that there was no correlation between amount read (language experience) and working memory span. Thus throwing some doubt on the role of language experience in syntactic processing. Finally, Caplan and Waters dispute the evidence supporting the idea that biological factors such as those suggested by MacDonald and Christiansen (e.g. aptitude of phonological distinction) could influence syntactic processing directly. The authors finish their rebuttal by reasserting the idea of modular, domain specific resources and cap the paper with the sentence: "Our view is that at the end of the day, domain-specific limitations on computational capacity will be left standing" (p. 73).

MacDonald and Christiansen respond to Caplan and Waters in another postscript. They do not target the specific concerns that Caplan and Waters raised, but rather focus their response on the sentence quoted above. They agree that capacity limitations exist and are important in defining the parameters of models of sentence processing, but dispute the nature of the limitation. For them, limitations will naturally arise through the limitations of spreading activation throughout the network, rather than having any predefined limits associated with any particular task or domain. Finally, they raise the concern that Caplan and Waters' bid to fractionate the verbal working memory system may lead to even more smaller and smaller fractionations, resulting in an overly restricted and complex system. They argue that this inherently conflicts with the interactive nature of both connectionist ideals, and sentence comprehension processes.

In the sections above we have discussed some of the various theories and proposals which relate to the interaction of working memory and language processing, and emphasised the question of domain specific versus domain general resource pools. In the next section, we shift our focus to examine the literature which attempts to investigate that issue using dual task methodologies.

## 1.4 Capacity Constraints and Dual-Task paradigms

As may have been inferred from this review so far, many studies have attempted to examine the architecture of working memory using differing methodologies. Some, as we have seen, have used individual differences (with High and Low span readers) to examine how working memory interacts with language processing (e.g. Just & Carpenter, 1992), while others have used dual task paradigms (e.g. Baddeley & Hitch, 1974). In the previous section we have, as part of our discussion regarding domain specificity, examined some of the individual differences literature, but in this section we will focus on what has been discovered using simultaneous dual tasks. The logic behind the dual task approach is that, in a limited capacity system, diverting resources to one task should interfere with the quality of processing of another. Should no interaction be found (assuming the central tenet of capacity constraint is correct), it implies that the two tasks draw on different pools of resources.

Many studies have used a variety of different secondary tasks to investigate the effect of stressing the memory system. While not all have had the goal of trying to answer the question posed by the Just and Carpenter vs. Waters and Caplan debate (are the resources available to the working memory system drawn from a single source (domain general) or are there multiple pools of specialised resources allocated for specific tasks (domain specific?)) many have been used as evidence to support one side or the other. The tasks have ranged greatly (see Conway & Kane, 2005, and Daneman & Merikle, 1996, for overviews) but can fit into a comparatively small set of categories. The first division is whether or not the secondary task interrupts the main task. In some studies (e.g. Wanner & Maratsos, 1978) the main task is interrupted by the presentation of the second task stimuli, whereas in others the two tasks are presented sequentially (e.g. task 1, task 2, resolution of task 1, resolution of task 2). The second division concerns whether or not the secondary task requires simple storage of information, entails processing, or entails both storage and processing components. Digit span (as well as letter span and word span) for example, entails only storage, whereas Daneman & Carpenter (1980)'s Reading span measure takes into account both storage and processing components. The final division comes from more recent research by Fedorenko, Gibson, & Rohde (2006) and Gordon, Hendrick, & Levine (2002) in which the relationship between the second task stimuli held in storage and the

stimuli in the main task is varied. Most secondary tasks involving only a storage component attempt to minimise semantic interference between the items to be stored and the items in the main task, however, the aforementioned researchers have shown that having the two stimuli streams interfere with each other in terms of lexical and semantic properties, can increase the effectiveness of a memory load significantly.

Much work has been conducted into which secondary tasks best interact with language processing and this formed the basis for much of Caplan & Waters (1999) arguments for a SLIR. We will briefly address the divisions discussed above in relation to the intention of our experiments.

#### 1.4.1 Interruptive vs. Non-Interruptive tasks

It has been shown that tasks which interrupt the processing of a sentence hinder both the processing of that sentence (e.g. Wanner & Maratsos, 1978) and the recall of the interruptive stimuli. This however, is likely to be less of a resource / capacity constraint issue, and more one of switching attention between two tasks, and the additional costs associated with task switching (as we have discussed in section 1.2.4.1). We therefore chose to present our extrinsic memory load sequentially, rather than interrupt the main task.

#### 1.4.2 Storage, Processing, or Both

Baddeley & Hitch (1974) found that maintaining six random digits in memory while performing an offline true / false judgement task influenced a participant's ability to reason accurately. They reasoned that the digits were being maintained in the phonological loop (storage buffer) which is structurally separate to the central executive. Daneman and Carpenter (1980) argued that digit span and other storage based span tasks are not adequate measures of working memory as they capture only one component of the system. As support for their claim they introduced the Reading Span task which, they argue, provides a much better measurement of working memory span, as it includes both language processing and storage components. We will discuss the Reading Span task in greater depth later in this section.

Just & Carpenter (1992) felt that their capacity constraint theory differed from the Baddeley & Hitch (1974) tripartite model in that storage and capacity are functionally rather than architecturally separable. In the tripartite model, all temporary storage was handled principally by the slave systems, and overseen by the central executive. In the multicomponent model, storage happens principally in the slave systems but is passed to the episodic buffer. The capacity constraint theory focused more on explaining the functioning of the part of the system that Baddeley and colleagues term the central executive, but accredited this area with storage abilities as well. The storage here, in Just and Carpenters view, represents the allocation of resources which are finite in capacity and split between storage and processing as is necessary. There exist, therefore, “trade-offs that are made between storage and processing” (Just & Carpenter, 1992. p. 124). It seems possible then, that with a sufficient storage load, processing will be detrimentally affected (this is the concept behind a supra-span presentation, see Vandierendonck, Vooght, & Van der Goten, 1998, for an example of its use). However, this trade-off idea has been refuted by Conway et al. (2005) who reports studies by other researchers which show that in offline measures of working memory, storage and processing are positively correlated. They argue that if storage and processing traded-off then they should see a negative correlation: as one increased the other should decrease. However, it should be noted that as Conway et al. point out, that participants are often directed to perform well on the processing task and are at ceiling levels of performance. The storage was not negatively affected but it is not clear from this evidence whether that is due to an excess of resources in those who reached ceiling level (i.e. the load was not sufficiently hard to tax their working memory systems), or if storage and processing do not compete for resources.

In their 1992 paper, Just and Carpenter do not make any clear statements concerning the existence of storage systems outwith those suggested by the capacity constrained theory. However, Just, Carpenter, & Keller (1996) consider the presence of additional storage areas to be functionally separate from those which they believe to trade resources with processing demands: “Several studies show, for example, that if a load is encoded and maintained in a peripheral buffer (i.e. the phonological buffer / articulatory loop), it does not interact with sentence complexity (Craik et al. 1990).” (Just, Carpenter, & Keller, 1996, p. 779). As we have noted before however, the studies of Craik, Morris, & Gick (1990) that Just et al. (1996) cite in support of this argument used self-paced reading as their measure of processing speed. We are concerned that this methodology is not sensitive enough to

pick up subtle online processing differences, and that Just et al.'s statement might be based on a lack of methodological fidelity. Just, Carpenter, & Woolley (1982) compared several different methodologies including self-paced reading (in both static and moving window forms) to online eye-tracking. They found mixed results but concluded that while in some ways self-paced reading provided qualitatively similar data to eye-tracking, in other ways it is qualitatively different (e.g. word skipping, preview etc.), and in many ways it is quantitatively different. Overall, Just et al. (1982) come down in support of self-paced reading: they argue that the results are similar enough to eye-tracking to allow researchers to continue using the paradigm rather than switching to the newer and more expensive eye-tracking system. However, some of the differences they observe (e.g. self-paced reading "appears to decrease the size of the word-length and word-frequency effects by a factor of two but to magnify most other effects by a factor of three or four" (Just et al., 1982, p. 233) appear, to us, to be more important than they consider them to be (but see also Mitchell & Green, 1978 for further discussion).

Daneman & Carpenter's (1980) Reading Span task has been widely accepted as an appropriate method of assessing working memory in participants and patients but it has not been without criticism. Concerns have been raised that it predicts working memory simply by invoking many cognitive processes, without specifying which components of the task are related to working memory (Daneman, 1982). Thus it leads to the problem that "one can argue that something analogous to working memory does appear to be an important factor, but we are left with few clues as to which of several factors may be of crucial importance." (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985, p. 120) but see also Daneman & Merikle (1996) for support of the Reading Span task.

The Reading Span task was designed as a measure of working memory span, rather than a method of reducing availability of resources during online processing. In fact it is difficult to see how one could tap the processing component that Daneman & Carpenter (1980) propose is critical, during an online study. We have considered various methods but in all cases, forcing participants to undertake extraneous processing while reading sentences, without interrupting the primary task, has proven infeasible. For example, if the secondary task was to perform a mathematical calculation which was presented prior to the main task sentence, participants would either perform the calculation and retain the answer during reading (storage only), or retain the equation and perform the calculation after having read

the sentence (also storage only). In addition, if you were to impose a time limit, either for the duration that the equation is displayed and / or the time allowed for response after completion of the sentence (so as to add time pressure and stop pre or post calculation), participants would interrupt the reading task, perform the mathematical calculation and then continue with the reading. As we have seen from the literature reviewed above, true division of attention is not impossible (e.g. Spelke et al., 1976), but is very hard to obtain and requires much time and exposure to the tasks.

The practicalities of adapting the Reading Span task to our needs are complicated, but there are other factors to consider as well. Essentially the Reading Span task involves two components: reading sentences (processing) and recall of the sentence final words (storage). In all likelihood, once those being tested become aware of the task requirements they will develop a strategy of using rehearsal or paying conscious and deliberate effort to the final words, in order to show aptitude on the test. In doing this they may reduce the processing they undertake on the sentence reading component. If it were possible to adapt this task to be used during an online study as a way of limiting available resources, these strategic factors would need to be taken into account.

One possible way to force an extrinsic concurrent processing load is to recruit processing demands from a secondary, non-verbal domain. Daneman & Merikle (1996) performed a meta-analysis of many different working memory measures and showed that, while less effective than domain specific tasks, aptitude on tasks from other domains (such as mathematical calculation) could predict ability in verbal comprehension. Conway et al. (2005) argue that this is because working memory resources are domain general, while certain skills are domain specific. Experts at particular tasks utilise domain specific skills to lighten the load on the domain general resources, but those with little experience with a task will rely more heavily on domain general resources. Therefore doing mathematical calculations draws on domain general resources, but will be a less efficient predictor of verbal comprehension because it does not take into account individual aptitude at either mathematics or verbal comprehension. We have already observed that the practicalities of integrating a mathematical task into an online study are complex, but it is possible that other domains might offer a way to tap processing resources.

Arrhythmic tapping during the main reading task may generate a processing load, and rhythmic tapping, at least, has been shown to interact with psycholinguistic processes such



as the phonological similarity effect (Saito, 1994). Other variations of the tapping task which have been shown to impact memory systems are Syncopated tapping (e.g. Larsen & Baddeley, 2003), Random Interval Repetition tapping (e.g. Vandierendonck et al., 1998), and matrix tapping (e.g. Vandierendonck, Kemps, Fastame, & Szmalec, 2004). However, none of these tapping tasks have a storage component which has been shown to be equally as important as processing (Daneman & Merikle, 1996). It is possible to conceive of a tri-task situation in which some form of tapping, a storage component such as digit span, and a main reading task were all present, but it is our opinion that this is problematic in two ways. Firstly, we are trying to avoid task switching costs and the disconnect between processing and storage (with them originating from two separate tasks) may cause a divide in attention and subsequently strategies for interleaving the two rather than running them concurrently. Secondly, the demands of keeping three tasks afloat at the same time may simply prove too difficult or too frustrating for participants.

#### 1.4.3 Semantic Interference in word list recall

Gordon et al. (2002) published a study which intended to impact the SR / SLIR debate. Using self-paced reading in a dual task paradigm, Gordon et al. manipulated the semantic relationship between a list of nouns and a to-be-read sentence, as well as the syntactic complexity of that sentence (subject-extract cleft vs. object-extracted cleft). The manipulation was essentially a word span task (storage only) but the items to be recalled either matched or miss-matched (in terms of semantic relationship) with the head nouns in the sentence. The authors found, in the accuracy of responses to comprehension questions, an interaction between syntactic complexity and semantic relationship in that the established deficit for processing object-extracted constructions vs. subject-extracted constructions was greater when the to-be-recalled words were semantically similar to the head nouns in the sentences, compared to when they were semantically unrelated. In the critical region (verb and head noun), for the self-paced reading times, there were also numerical trends of an interaction between complexity and semantic similarity in the same direction as the comprehension question finding. The authors took this as online support for the comprehension question finding. They specify that cue-retrieval is the site of interference as it is the similarity between the items being held in memory and the items being constructed into the discourse which causes the interference. Interestingly they found this to support the SR approach rather than the SLIR approach as it shows

interactions between systems which Waters and Caplan would define as interpretive and post-interpretive, and thus would be supported by separate resource pools. These results are replicated and extended by Fedorenko et al. (2006). Fedorenko et al. showed a significant online result (self-paced reading) which was the same as Gordon et al.'s numerical trend (i.e. an interaction between syntactic complexity and similarity of memory nouns to nouns within the sentence, such that the critical region was read slowest in the object-extracted / similar nouns condition). Fedorenko et al.'s experiment differed slightly from Gordon et al.'s in three ways: Firstly, to increase naturalness, the sentence was presented in a moving window rather than centrally. Secondly, while Gordon et al. used cleft subject / object extracted sentences, Fedorenko et al. did not use clefts. Thirdly, Fedorenko et al. also manipulated the number of memory nouns which a participant was required to maintain. This final manipulation allowed Fedorenko et al. to show that, even when utilising the interference generated from similar memory nouns, the amount of load (i.e. the number of items held in memory) is still critical to observing memory load effects. With only one memory noun, no interaction was observed, but with three memory nouns, the interaction mentioned above was present.

Following Gordon et al., Fedorenko et al. explain their results in terms of interference during retrieval of memory items. They argue that during subject-extracted sentences each new sentence item is adjacent to the one it needs to attach to, and so can be immediately attached without requiring an appeal to memory. However, in object-extracted constructions, this is not the case. Nouns require attachment to an item several words previous, and so an appeal must be made to memory in order for that attachment to occur. The retrieval of the to-be-attached-to noun held in memory experiences interference if the semantic properties of other items being held in memory overlap or are very similar (see Van Dyke & McElree, 2006, for support of this account). Another possibility which Fedorenko et al. highlight (but subsequently dismiss) is that rather than the site of interference being retrieval; it may well be that syntactic storage costs (at the encoding and maintenance stages) are at the root of the conflict. It is possible, they concede, that having multiple items held in memory which share very similar properties and features, inflates the storage costs of those items to the point where other processes (like the processing of the main task sentence) suffer. This account finds support in work by Chen, Gibson, & Wolf (2005), but is not favoured by Fedorenko et al. who call on the work of Gardiner, Craik, & Birtwistle (1972) to support their retrieval based explanation.

The authors of both studies argue that their findings support the single resource theory of Just and Carpenter (1992) but we are unconvinced that what they show here impacts directly on that debate. It is clear that semantic interference occurs between items held in memory and on-going sentential processing, and on the surface this would appear to belie the defining partition of the SLIR hypothesis: the divide between interpretive and post-interpretive processes. However, the nature of the interference needs to be carefully considered. Critics of the capacity constraint approach (e.g. Navon, 1984) have highlighted the undefined nature of what constitutes 'capacity' or 'resources'. This lack of specificity makes it hard to pin down what constitutes a resource conflict, and what might be accredited to other factors. In this case, our feeling is that semantic interference between memory traces falls outwith the concept of resources as it was intended by Just and Carpenter (1992) and Waters and Caplan (1996). It is true that the results described above do indicate an interaction between syntactic processing and extrinsic memory load, but it is not, in our opinion, the competition for resources which is causing the interaction. More appropriate to our conception of the meaning of 'resources' is the possibility that it is the cost of encoding items with similar semantic properties which is depleting resources, rather than semantic interference at the retrieval stage. However, even this adds the additional step of requiring the items to have similar semantic attributes. If this was truly a resource issue then presumably the effect observed by creating additional load in this way could be replicated simply by increasing the number of items held in memory. The point being that if it is a pure resources conflict, interference would be a means of generating the level of load required to observe an effect, rather than a necessary and indispensable component. While Fedorenko et al. (2006) do conclude that both the number of items held in memory, and semantic interference are necessary to generate the observed effect (the interaction disappears when only one item is held in memory, regardless of whether it provides semantic interference or not); they demonstrate that a memory load of three items is not enough to do so on its own. Fedorenko et al. also state that previous studies (including many by Waters and Caplan, and colleagues) have failed to find interactions between interpretive and non-interpretive processes while using digit span, because of the lack of semantic interference provided by the stored representations of digits. They argue that this was a contributing factor to Caplan and Waters proposing domain specific resource pools.

We feel that the assertion that interference is key to observing resource conflict is sidestepping the core debate by adding additional factors into what is a fairly binary

distinction. We do not seek to unfairly disregard the findings of Gordon et al. and Fedorenko et al., and fully agree with their interpretation of their results, but we are concerned that the semantic interference account obfuscates the SR / SLIR debate, and it is for this reason that we chose not to utilise it in our research.

#### 1.4.4 Digit Span

As we have discussed above, digit span is now considered a test of short term memory, rather than working memory but, to our knowledge, no test of working memory (i.e. containing both storage and processing components) exists which can be adapted to the task of reducing processing resources during online sentence comprehension. It is hard to conceive of such a non-interruptive task in which attention is divided between concurrent processes without considerable time and training.

It is worth noting, however, that claims made about the lack of interaction between storage only tasks and sentence processing have been made exclusively from offline and partially offline data. This data has been a mix of reasoning tasks (e.g. Baddeley & Hitch, 1974), truth verification latencies (e.g. Craik et al., 1990), and self-paced reading (e.g. Craik et al., 1990; Daneman & Carpenter, 1980; Just & Carpenter, 1992). Both the authors of the multicomponent model and the capacity constraint theory would anticipate some trade-off between processing and storage, whether it be in terms of the involvement of the central executive or episodic buffer in providing overflow storage, or the functional rather than structural separation of resource expenditure. With that in mind it seems logical that if we place participants under an extremely high extrinsic memory load, even if that load is storage only, it should influence the availability of resources for processing tasks. It is possible then, that the studies which have not found interactions between sentence processing and digit span load have been deficient in either the amount of load applied, or the fidelity of their tools to observe the interaction.

As mentioned above, Gordon et al. (2002) and Fedorenko et al. (2006) concluded that semantic interference between memory items and sentence items is necessary to show interactions between sentence processing and memory loads. However, both studies used self-paced reading (albeit one with a moving window rather than central presentation) and neither study exceeded a memory load of three items. If there was an interaction between

sentence complexity and memory load to be found, the controls in these experiments were not likely to find it.

### 1.5 Overview of the current study

Our intent is to use eye-tracking to investigate the role of working memory in sentence processing. Specifically we will be using a high extrinsic memory load (digit span) to investigate the role that the availability of cognitive resources plays in the contextual influence on syntactic and lexical disambiguation, and semantic prediction. Our principal goal is to provide evidence impacting the debate between Domain General (Single Resource (SR theory) - Just & Carpenter, 1992) and Domain Specific resources (Separate Language Interpretation Resource (SLIR theory) - Caplan & Waters, 1999). In the process of examining the effects of resource reduction we will also observe normal (non-loaded) processing independently, and will report any findings of interest.

The literature review we have provided above focuses on the literature pertaining to Working Memory and resource allocation. Each of the upcoming individual experimental sections contains a further brief literature review which is directly relevant to the topic being addressed.

In our experiments we will titrate the digits presented as part of the digit span task to generate an appropriately high level of load for each participant. The next chapter in this section provides details of how the digit span task was created, titrated, and applied to each participant.

In Section 2 we present two experiments and a meta-analysis in which we examine how being under load influences the use of a preceding context when encountering and resolving syntactic ambiguities.

Section 3 reports an online experiment designed to guide the creation of materials for experiments examining load and context processes for lexical ambiguity. Subsequently, three experiments and a meta-analysis are reported.

Section 4 details a single experiment in which we make an initial investigation into the interactions of Load, context, and semantic prediction.

Overall findings and conclusions will be discussed in Section 5.

## Chapter 2: Digit Span Task

### 2.1 Introduction

As discussed at the end of the previous chapter we use a titrated Digit Span task to implement an appropriately high level of loading which is tailored to each participant's individual working memory span. Following the work by Crannell & Parrish (1957), Miller (1956) and Baddeley & Hitch (1974) we centred the number of presented digits at six per digit string, and allowed a variance of three. Therefore the minimum number of presented digits per string was three, and the maximum was nine.

One concern which we wanted to address was the possibility that certain digit strings could have an undue influence on the availability of working memory resources due to properties of those specific digits. For example, it is reasonable to assume that a string containing many syllables (seven, zero, seven, three, two) might require more resources to store and rehearse than a string with fewer (nine, one, six, three, two). For this reason a computer program was created to generate fully randomised sets of digits which could then be associated with the experimental sentences, with each sentence seeing a different set of digits in each condition, across all levels of loading, and for each participant.

Rules of Digit Strings:

Several rules were imposed upon the Digit Generation program in order to stop particularly easy or difficult strings appearing.

1). No ascenders or descenders: We did not allow numbers to appear adjacently if they were a plus or minus one increment of each other (e.g. 19568 would not be allowed, nor would 19548).

2). No local repetitions: We did not allow numbers to appear twice in the same string (e.g. 19579 would not be allowed)

3). No global repetitions: We did not allow the same string of digits to appear twice to the same participant. The concern here was that, due to the randomisation of the presentation software, we might encounter two sequential items with the same digits, causing significantly easier recall for the second trial in the sequence.

## 2.2 Creation of Digit Span Task

Digit strings were created using the computer program mentioned above and outputted as CSV files. These were then imported into Microsoft Excel and collated with the experimental items and fillers. Each set of items was then output as a single list and labelled by the number of digits in the digit string. 32 lists were created for each of the seven digit string possibilities (3-9) and were used incrementally per participant to maintain full randomisation.

## 2.3 Digit Span Pre-Test and Titration

The number of digits in the digit string which was presented to each participant varied depending on their performance on a brief pre-test. The pre-test was conducted using sentences similar in length and structure to the experimental items, and started with a digit string of 6. Using a specially written program which simulated the experiment, participants were instructed memorise the digits and then press a button to display the sentence. They had to read the sentence silently and press a button when finished. No time limit was placed on either the digit encoding or sentence comprehension sections. Participants were then asked by the program to say the digits out loud, and the experimenter compared the responses to a list. A comprehension question followed. Randomisation did not occur in the pre-test so the experimenter's list of digits and comprehension questions was pre-prepared. Responses were considered correct only if the digits, the order they were recalled in, and the answer to the comprehension question were accurate. If correct the program presented another digit / sentence item increasing the digit string by one digit. If incorrect the participant was given the choice to try the same level again, or reduce the digit string by one. Participants repeated the process until they were able to give correct responses twice in a row at a certain level of loading, but not able to do so at the next level up.

Once the appropriate level of loading had been established, the main body of the experiment commenced. The procedure was almost identical to that conducted in the pre-test. The exceptions were that a). A sixty second time limit was imposed by the software on the digit encoding and sentence reading sections (although participants were encouraged to try to spend only around 10-15 seconds encoding the digits, and the sixty second time-

out was never reached on either component), and b). No feedback was given on the digit recall (except on the practice trials).

It was decided to structure the experimental procedure to be: 1). digit presentation, 2). sentence presentation, 3). digit recall, 4). comprehension question, in order to try and homogenise the amount of time between digit presentation and recall across trials and participants. Had we structured it so that the comprehension question preceded the digit recall there would have been greater temporal variability between the digit components.



## Section 2: Cognitive Load, Context, and Syntactic Ambiguity Resolution

Language, by the nature of its incremental delivery, is inherently temporarily ambiguous. In this section we will be focusing on Syntactic Ambiguity which is also known as Structural Ambiguity. A classic example of this comes from Bever (1970):

*The horse raced past the barn fell.*

People encountering this sentence tend to read it as ‘The horse raced past the barn’, and then find they have an additional word (fell) which cannot be attached to the syntactic structure they have created. The difficulty here is that the verb *raced* can be active or passive and the attachment of constituents downstream depends on which way it is interpreted. There is a general bias to interpret it as active (interpreting it as a main clause verb) rather than as passive (interpreting it as the onset of a reduced relative clause) which leads people away from the correct interpretation. The example below (also taken from Bever (1970), emphasis added) disambiguates the verb voice allowing only the correct interpretation.

*The horse that was raced past the barn fell.*

Critically, the ambiguity is not enduring. Once a certain point of the incrementally delivered input is reached, the comprehender becomes aware of the ambiguity and, if they have been misled by it, must engage in reanalysis. Sentences which entail this process of committing to a structural analysis which later, when more information becomes available turns out to have been incorrect, are *garden-path* sentences as comprehenders feel they have been led down the garden path.

Many different forms of garden path sentences exist, but further examples of the type displayed above are given by Farmer, Cargill, & Spivey (2007) who provide an excellent review of syntactic ambiguity and the models proposed to explain analysis, reanalysis, and recovery. We will briefly summarise those models here before moving on to consider how syntactic ambiguity, context, and cognitive loading interact.

Broadly speaking there are three types of model which seek to explain how a comprehender handles syntactic ambiguities like the one presented above:

Firstly, coming from a modular perspective (Fodor, 1983), there are serial models which posit that syntax has a privileged status when dealing with ambiguities (Clifton & Ferreira, 1989; Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Rayner, Carlson, & Frazier, 1983). There is no immediate influence of other factors such as context or semantic plausibility and the most syntactically likely interpretation of the ambiguity is automatically accepted until further information becomes available and causes the comprehender to question its validity, at which point reanalysis occurs.

Secondly, there are parallel competition models in which multiple representations are created every time a syntactic ambiguity is encountered (Bates & MacWhinney, 1989; Crain & Steedman, 1985; MacDonald et al., 1994; McClelland & Rumelhart, 1986; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Rumelhart & McClelland, 1986; Tabossi et al., 1994). The representations then compete with and inhibit each other in order to achieve enough activation to be selected.

The third type of model is called an unrestricted race model, and contains elements of both the serial and competition models (Traxler, Pickering, & Clifton, 1998; van Gompel, Pickering, Pearson, & Liversedge, 2005; van Gompel, Pickering, & Traxler, 2001). The unrestricted race model allows multiple representations to be created whenever an ambiguity is observed, but those representations do not compete with each other, instead they operate independently and whichever gains sufficient activation for selection first, is then selected. The other representation(s) is then inhibited.

In this section we are interested in examining how the limitation of cognitive resources interacts with both syntactic ambiguity, and the contextual disambiguation of syntactic ambiguity. We will first discuss the main principles behind Capacity Constrained models of language processing and syntactic ambiguity, then briefly summarise the research pertaining to the role that prior context plays in disambiguating syntactic ambiguities. We will then introduce a series of experiments investigating how the use of context to disambiguate syntactic ambiguities is affected by the limitation of cognitive resources.

Spanning the models of sentence processing described above is the concept of a resource constrained system which functions faster and more accurately when there are more resources available (Just & Carpenter, 1992; King & Just, 1991; MacDonald, Just, & Carpenter, 1992).

Although not dealing with syntactic ambiguity directly, King & Just (1991) reported two experiments relating to the availability of working memory resources, and contextual facilitation. They took participants of low and high reading spans (as defined by the Reading Span task discussed in sections 1.3 and 1.3.1 of the literature review) and had them read Subject relative and Object relative sentences like those below.

**Subject Relative:**        *The reporter that attacked the senator admitted the error.*

**Object Relative:**        *The reporter that the senator attacked admitted the error.*

With reference to several other bodies of research, King & Just (1991) argue that the processing of an object relative syntactic construction taxes the working memory system more intensely than a subject relative construction does. They predict an increase in the self-paced reading times of the critical regions (the end of the relative clause (attacked or senator) and the main verb (admitted)), and the comprehension accuracy would be negatively affected (with, for example, participants concluding incorrectly that the senator admitted the error).

In the first experiment they manipulated cognitive load by performing a pseudo online version of the reading span task. Trials varied from having only one sentence to having three, with the target sentence as the final sentence. At the end of the trial participants had to recall the final word of each sentence they had read. They could, therefore, be retaining one or two sentence final words at the point of reading the target sentence, and would have to recall one, two, or three sentence final words at the end of the trial. The overall findings of this experiment were that low span and high span readers performed similarly when reading subject relatives, but that the comprehension and reading times of low span readers were significantly poorer and slower when reading object relatives, while high span readers seemed to suffer much less. With the introduction of load, no reading time deficits were observed between span groups, but a deeper analysis combining reading times and

comprehension results revealed that low span readers were not spending enough time on the critical regions to generate the correct interpretation of the sentence. They were, instead, rushing through the reading components in a strategy designed to increase accuracy on the load recall, which, unfortunately meant their comprehension suffered and their reading times were not indicative of the time it would actually take them to generate a correct parse.

King and Just, surprised and concerned at the very low comprehension rates of the low span readers, hypothesised that in normal reading circumstances there are generally contextual clues to support this type processing which were lacking from the sentences in the first experiment. The second experiment was designed, therefore, to see if low span readers were able to make use of these cues to reduce their comprehension deficit.

Using sentences of the object relative construction, but with pragmatically guiding verbs, King and Just manipulated cues as to which of the two actors in the sentence was the agent of the given verb. As can be seen in the examples given below, the verbs vary to create cohesion with the actors. This, King and Just argue, should reduce comprehension errors in which low span participants accredit the wrong actions to the actors, but that it may not be observable as differences in the reading times as the contextual information is not presented in time to influence the syntactic processing of the object relative.

**Fully Supportive:**        *The robber that the fireman rescued stole the jewellery.*

**Supportive of Relative:** *The robber that the fireman rescued watched the program.*

**Supportive of Main:**    *The robber that the fireman detested stole the jewellery.*

**Unsupportive:**         *The robber that the fireman detested watched the program.*

The principal findings of experiment two were that pragmatic cues did indeed significantly improve the comprehension of object relatives, especially for the low span readers whose comprehension without them was particularly poor. As predicted, the addition of pragmatic cues did not strongly influence reading times as they occurred after the bulk of the syntactic processing (which was the cause of the observed slow down) had been completed.

Following on from King and Just, MacDonald et al. (1992) conducted three experiments focusing specifically on how participants with low and high reading spans processed syntactic ambiguities. They proposed and tested a model called the Capacity Constrained Parsing Model which relates specifically to how the processing of temporal syntactic ambiguities is influenced by the availability of cognitive resources. Macdonald et al. take a parallel competition model (of the type discussed briefly above) as the parsing system on which to base their capacity constrained system.

In a parallel competition parsing model, each time a point of syntactic ambiguity is reached multiple representations relating to the different possible syntactic interpretations are created. These representations vie for selection by accruing activation which is gained from various properties of the text, as well as probabilistic sources (how frequently a particular grammatical structure occurs).

The capacity constrained model predicts that the duration that each of these representations is maintained for is reduced in people who have fewer working memory resources available (e.g. those with a lower reading span). Therefore a low span reader will generate multiple representations equally as well as a high span reader, but they will abandon much more quickly those representations which do not gain much activation. Only the most highly active representation is maintained and so, should that representation turn out to be incorrect, a low span reader suffers much more heavily than would a high span reader who is still maintaining that representation. The model predicts, therefore, that in the case of a garden path in which the preferred representation is shown to be incorrect, a low span reader will have poorer comprehension or longer reading times as the representation will have decayed and be irretrievable. They will either fail to parse the sentence correctly, or will have to engage in full reanalysis and repair. Conversely, high span readers under the same circumstances will have better comprehension and will not have to engage in full reanalysis as the representation for the correct interpretation is still available. Conversely however, the maintenance of multiple representations taxes the high span reader and reduces the resources they have available to conduct other processes, so in the situation in which the preferred interpretation turns out to be correct and no reanalysis is required, high span readers may actually have longer overall reading times than low span readers.

In their first experiment, MacDonald et al. used unambiguous sentences, and comparable reduced and non-reduced relative clause constructions like the ones given below to measure self-paced reading times and comprehension accuracy of high, medium, and low span readers. The argument is that because the reduced relative construction is fairly rare and complex it differs sharply from the common and straightforward main verb construction. The representation for the relative clause interpretation, therefore, should be activated less strongly, and should decay faster than that of the main verb construction.

- a. The experienced soldiers spoke about the dangers before the midnight raid.

Main Verb Interpretation, Unambiguous

- b. The experienced soldiers warned about the dangers before the midnight raid.

Main Verb Interpretation, Temporarily ambiguous

- c. The experienced soldiers who were told about the dangers conducted the midnight raid.

Relative Clause Interpretation, Unambiguous

- d. The experienced soldiers warned about the dangers conducted the midnight raid.

Relative Clause Interpretation, Temporarily ambiguous

In Sentences a. and b. the verb following *soldiers* is resolved in favour of a main verb interpretation. Sentence a. is not ambiguous due to the nature of the verb *spoke*, whereas in sentence b. *warned* could be either the main verb, or the introduction to a relative clause, and participants are not able to select which representation is correct until the end of the sentence. Sentences c. and d. resolve in favour of a relative clause interpretation, with c. being unambiguous due to the presence of the relativizer *who were*, and d. being ambiguous until the verb *conducted* occurs.

MacDonald et al. predicted that when the ambiguous sentence resolved in favour of the main verb construction, because low span readers were only maintaining the correct representation they would have faster reading times than high span readers who were

maintaining both representations and delaying higher level processing until the ambiguity was resolved and the resources it tied up became available. The model also predicts that longer reading times should occur for all reading span groups when the relative clause interpretation was correct, as higher span readers maintained multiple representations, and lower span readers had to engage in reanalysis. They do not however, make predictions about the severity of the slow down for each group, or if the two different sources of slow down would be quantitatively different.

Their findings supported their predictions, to summarise:

For main verb resolutions all readers showed longer reading times for ambiguous sentences, but higher span readers showed greater effects of ambiguity than did lower span readers. This reading time difference was most highly observable in the disambiguating region. High span readers also had longer reading times in the post-ambiguity regions than low span readers did when an ambiguity was present.

For relative clause resolutions the pattern is less clear as all reading span groups struggled with correctly comprehending the sentences, and the low span group performed significantly more poorly on the comprehension questions than the medium or higher span groups. This led to complications in the processing of reading times as, if the low span reader did not comprehend the sentence correctly, they will have spent less time on the critical disambiguating regions than other low span readers who did. Subsequently, for the purposes of interpreting the reading time data, two smaller sub-groups appear within the low span group causing issues of power within the experiment. Overall however, reading times did show an effect of ambiguity but it did not interact significantly with reading span group. MacDonald et al. rationalise that the reading times were not significantly different due to each group's reading times increasing because of different mechanisms. The high span group had higher reading times due to maintaining multiple representations and selecting a dispreferred one, while the lower span group showed longer times based on reanalysis and repair procedures. These mechanisms are not distinguishable from the data here.

MacDonald et al. subsequently reported two further experiments, the second of which is not particularly relevant to this thesis so shall not be discussed. The second experiment

however, shows that the lexico-syntactic properties of nouns can have a direct and immediate influence on resolving this type of syntactic ambiguity.

Concerned that the verbs used in the previous experiment might be creating sources of difficulty in their own right (transitive verbs having more argument options than intransitive verbs), they set out to distinguish between difficulty added by allowing additional argument structures, and allowing relative clauses. To do this they replaced half of the common nouns in the sentences from the previous experiment (*the experienced soldiers*) with proper nouns (*Colonel Wilson*) as proper nouns do not allow reduced relative constructions to follow them directly. They found increased reading times for those sentences using common nouns, but not for those using proper nouns. They interpret this as a replication of the previous experiment, with the extension that rules out the ambiguity of the verb as a source of difficulty. In terms of this thesis this is relevant as it is an early demonstration that lexico-syntactic properties can influence syntactic ambiguity resolution, and, as will be discussed in more depth shortly, our contextual manipulation involves the lexico-syntactic properties of referential context.

The models and experiments discussed above show how cognitive resource constraints can impact syntactic ambiguity resolution, we have covered the main principles but there have been many further investigations in this area (e.g. MacDonald et al., 1994; MacDonald, 1993; Pearlmutter & MacDonald, 1992; Pearlmutter & MacDonald, 1995) but see also Vos & Friederici (2003) for ERP work countering the general support for parallel competition models seen in these papers.

#### *Prior Context, Referential Contrast Sets, and Syntactic Ambiguity*

In 1988 Altmann & Steedman published a paper reporting an experiment the results of which countered the then popular idea of a modular, syntax first parser (one of the principal serial models mentioned above). The experiment manipulated a context sentence to bias the reading of a subsequent target sentence which contained a noun-phrase / verb-phrase ambiguity.

In the example given below, the context sentences (a.) and (b.) define a contrast set of either two safes or a safe and a strongbox. The target sentences (c.) and (d.) contain a



prepositional phrase (initiated by the word *with*) which either attaches to a noun-phrase (which safe? The safe with a new lock) or to a verb-phrase (how? With the dynamite). By combining context a. with target c. you have a referentially supported noun-phrase attachment, because specifying which safe is blown open is necessary for full comprehension. Combining a. with d. on the other hand, leaves the reader unsure of which safe was raided. Context b. combined with target c. is redundant as there is only one safe, while combining b. with d. provides a referentially supported verb-phrase attachment as the word *safe* clearly selects the appropriate noun, and under the Minimal Attachment hypothesis (Frazier, 1978; Frazier & Fodor, 1978), the VP should be the selected interpretation by default. Interestingly, in order to investigate whether the redundancy of the VP attachment (why specify that the burglar used dynamite if that is obvious from the context?), would influence reading times, Altmann and Steedman created another condition in which there was a choice of implements (...carrying some dynamite and some gelignite).

- a. A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a safe with an old lock.
- b. A burglar broke into a bank carrying some dynamite. He planned to blow open a safe. Once inside he saw that there was a safe with a new lock and a strongbox with an old lock.
- c. The burglar blew open the safe with the new lock and made off with the loot.
- d. The burglar blew open the safe with the dynamite and made off with the loot.

The experiment was therefore a 2x2x2 crossing the NP / VP supportive contexts, with the NP / VP attached targets, and the one / two implement(s). Two versions of the experiment were run, one using whole sentence self-paced reading, and one using phrasal self-paced reading. The results patterned the same way in both versions.

The one / two implements manipulation made no difference at all. The overall result of note however, was that sentences with contexts which supported the target attachment had lower reading times than did those which did not. This neatly showed that an appropriate prior referential context can facilitate the processing of a syntactic ambiguity online.

Several other studies have expanded on this finding, its generalizability to other forms of syntactic ambiguity, and its implications for models of parsing (e.g. Altmann et al., 1994; Altmann, Garnham, & Dennis, 1992; Britt, 1994; Farmer et al., 2007; Mohamed & Clifton, 2011; Spivey & Tanenhaus, 1998)

Binder, Duffy, & Rayner (2001) and Traxler & Tooley (2007) in particular have extended what is known of context effects on parsing to include reduced relative clause ambiguities such as those used by King & Just (1991) and McRae, Spivey-Knowlton, & Tanenhaus (1998). It is clear from this body of research that reduced relative clauses are hugely biased towards a main clause interpretation but that this bias is susceptible to contextual influence. Traxler & Tooley (2007) used sentences such as those presented below to investigate the influence referential contrast sets on the garden path's created by reduced relative clause processing. The parentheses indicate the conditional presence of the relativizer *who was*.

- a. The students were looking for someone to speak at the awards dinner. They interviewed a possible speaker.

*(One referent context)*

- b. The students and the teachers were looking for someone to speak at the awards dinner. They each interviewed a number of possible speakers.

*(Two referent context)*

- c. The speaker (who was) selected by the students would work perfectly for the program.

*(Target sentence)*

In the example above, the verb 'selected' can be interpreted as the onset of a main clause, or as part of a relative clause. The sentence is temporarily ambiguous and is disambiguated

in one of two ways. Either the relativizer disambiguates the sentence in favour of the relative clause interpretation when it is present or, when it is not present, the word 'by' indicates to the reader that the main clause interpretation cannot be correct (and so disambiguates in favour of the relative clause). Traxler and Tooley argue that in this case, the single referent context does not favour either the main clause interpretation or the relative clause interpretation, whereas the multiple referent context (referentially at least) goes against the normal syntactic bias to favour the reduced relative structure. The idea behind Traxler & Tooley's experiment was that the single referent context would not provide as much support for the reduced relative interpretation as the multiple referent context, and so greater difficulty (manifesting in longer reading times) should be observed on the disambiguating 'by' region in the case of single referent context. Thus they created a 2x2 experiment crossing ambiguity (based on the presence or absence of the relativizer) and contextual support (based on the single or multiple referential contexts).

Using four measures of eye-tracking (First Pass, Regressions Out, Regression Path, and Total Time), they looked for evidence that contextual information could override, or at least ease the recovery from, the garden path. They found that supportive referential contexts could not override the natural bias to interpret the reduced relative as the onset of a main clause, but that they did reduce the impact of the garden path.

In the next section we will report a series of experiments in which we will make use of this strongly unbalanced form of ambiguity, and established form of contextual mitigation, to highlight how contextual effects might be influenced by a reduction in cognitive resources.

## **Chapter 3: NoLoad, Context, and Syntactic Ambiguity (Experiment 1)**

### **3.1 Introduction**

This is the first of two experiments examining how context and syntactic ambiguity interact while free from or under cognitive load. As we have seen from the literature discussed above it is known that non-syntactic factors can influence the selection of syntactic representations when points of ambiguity are being processed. The question of how syntactic processing is influenced by the availability of cognitive resources has, to our knowledge, only been investigated using a combination of predefined high and low span groups (an individual differences approach), partially offline reading measures (self-paced reading), and offline comprehension questions. We aim to experimentally manipulate the availability of resources by using a dual task paradigm with the secondary task involving a high extrinsic memory load titrated to each participant. In this way we intend to ensure that all participants have the same resources available (and to vary that amount), regardless of whether they might be part of a high, medium, or low span group as a base level. We will also be using fully online eye-tracking to allow examination of the full scope of reading behaviours (for example, self-paced reading does not allow for regressions or re-readings). We will be using offline comprehension questions but critically, while MacDonald et al. (1992) used comprehension questions as a critical measure (and as such ran into problems with the analysis of reading time data from the low span group some of which correctly interpreted the sentence and some of which did not), we use it only to check that comprehension has occurred correctly. Our intention is to analyse reading time data only from trials where comprehension was successful. This will allow us to examine the processes of disambiguation without the confound of including data from trials in which the ambiguity was not noticed and resolved.

In the experiments presented in this chapter we follow from the work of Binder et al., 2001; King & Just, 1991; McRae et al., 1998; Traxler & Tooley, 2007 and others in using reduced relative clauses to provide a syntactic ambiguity, and referential contrast sets as contextual information to facilitate their resolution.

In the example below, the underlined sections display the contextual reference set changes between neutral and supportive contexts, while the presence / absence of a relativizer (indicated by italicised parentheses) provides the reduced relative clause ambiguity.

The Smith twins, still only babies, are no strangers to excitement. Only last week, a freak storm caused a fire which frightened one of the babies but not the other. The baby (*who was*) frightened by the flames was crawling away as quickly as she could. Her mother picked her up and tried to calm her down.

The Smith family's baby and 7-year-old are no strangers to excitement. Only last week, a freak storm caused a fire which frightened the baby but not the 7-year-old. The baby (*who was*) frightened by the flames was crawling away as quickly as she could. Her mother picked her up and tried to calm her down.

In the first experiment we do not include a resource manipulation. This experiment will examine the online effects of contextual facilitation of syntactic ambiguity and will provide a baseline for our lab and materials against which the subsequent resource manipulation experiment will be compared.

In brief, in this non-resource manipulated experiment we cross ambiguous and unambiguous sentences with supportive and unsupportive contexts. Given the models detailed above we might expect to see evidence of:

A main effect of ambiguity such that ambiguous sentences incur longer reading times in the disambiguating region than unambiguous sentences. This would provide support for syntax first models as it would show that ambiguity and context were not interacting.

An interaction such that difficulty is caused by the ambiguity if the context is not supportive, i.e. the Garden Path effect (caused by the presence of an ambiguity) should be greater without contextual support. Therefore we expect to see longer reading times on (and possibly immediately after) the disambiguating region for ambiguous sentences when the context is unsupportive compared to when it is supportive. Unambiguous sentences should pattern the same but the effect of contextual facilitation will be smaller.

## 3.2 Methods

### *Participants*

36 participants meeting the following experimental criteria took part in the study.

All were Native speakers of British English, over 18 years of age, had no known reading disorders (e.g. dyslexia), and had normal or corrected to normal vision.

Participants were required to correctly answer 80% or more of the comprehension questions or their data was discarded. For this experiment three participants were excluded (and others run to replace them) on the basis of failure to meet the comprehensions question threshold or poor tracking. The range for the Comprehension Questions of those included was 85% - 98%, and the Mean was 92%.

### *Stimuli*

The 28 experimental stimuli were created as follows. Each item begins with two context sentences, the first of which either distinguishes two referents (e.g. a baby and a 7-year-old) or uses one term to refer to multiple referents (e.g. the babies), while the second implicates one of the referents as the subject of the next sentence. The target sentence (derived from McRae et al., 1998 and developed for Dubey, Sturt, Keller, & Stewart, in prep) follows, and consists of a noun phrase selecting one of the referents (NP1), a verb (V), a second noun phrase (NP2), and some spill over material for discourse cohesion. Finally, a fourth sentence is included to round off the discourse. Critically, between NP1 and the Verb in the target sentence there is either a reduced relative clause or non-reduced relative clause. The difference being the presence or absence of the 'who was' relativizer.

The presence of a relativizer removes the syntactic ambiguity which is otherwise present; without it readers are naturally biased to interpret NP1 as the agent of the sentence, rather than the patient. The context either facilitates the correct interpretation of agency by causing readers to expect further information distinguishing between referents, or implies the incorrect interpretation in the case where such distinguishing information is already present. We therefore have a 2x2 design with Context (Supportive / Unsupportive) being crossed with Ambiguity (Ambiguous / Unambiguous).

A set of example stimuli is given in 3.2.a - 3.2.d. See Appendix 1 for a full list.

3.2.a.

The Smith twins, still only babies, are no strangers to excitement. Only last week, a freak storm caused a fire which frightened one of the babies but not the other. The *baby frightened* by the flames was crawling away as quickly as she could. Her mother picked her up and tried to calm her down.

(Supportive Ambiguous)

3.2.b.

The Smith twins, still only babies, are no strangers to excitement. Only last week, a freak storm caused a fire which frightened one of the babies but not the other. The *baby who was frightened* by the flames was crawling away as quickly as she could. Her mother picked her up and tried to calm her down.

(Supportive Unambiguous)

3.2.c.

The Smith family's baby and 7-year-old are no strangers to excitement. Only last week, a freak storm caused a fire which frightened the baby but not the 7-year-old. The *baby frightened* by the flames was crawling away as quickly as she could. Her mother picked her up and tried to calm her down.

(Unsupportive Ambiguous)

3.2.d.

The Smith family's baby and 7-year-old are no strangers to excitement. Only last week, a freak storm caused a fire which frightened the baby but not the 7-year-old. The *baby who was frightened* by the flames was crawling away as quickly as she could. Her mother picked her up and tried to calm her down.

(Unsupportive Unambiguous)

*Procedure*

The 112 experimental materials (28 items in 4 conditions each) were divided into 4 lists (Latin Square design). The experimental items were presented in a pseudo randomised order such that each participant saw only one condition of each item. A total of 72 fillers (of varying sentence construction and overall length) were included in the randomisation. The randomisation disallowed two experimental items to occur consecutively and made sure that a filler always intervened.

The experiment was run using an Eyelink 1000 desk mounted eye-tracker, with a sampling rate of 1000hz. Right eye only was tracked. The system was calibrated and if calibration was successful the experiment proceeded. Items were presented using software developed at the University of Massachusetts at Amherst, called Eyetrack<sup>1</sup>. Each trial consisted of three components (a gaze contingent trigger, a reading component, and a comprehension question). The gaze contingent trigger was used to verify calibration. It took the form of a black square vertically centred but offset to the left side of the screen. When this square was fixated by the participant it was automatically replaced by the reading component. The reading component consisted of an experimental item or filler. The participant pressed a button on a game pad when they had finished the reading component. Doing so moved them on to third component: a comprehension question. Comprehension questions were yes / no questions about a non-critical region of the preceding item, and occurred half of the time. The remaining half was filled with a statement saying that there was no question about the preceding sentence (this was done to allow later experiments involving load to maintain more similar timings between digit onset and recall). Questions and statements were responded to using pre-specified buttons on the game pad which corresponded spatially to prompts on the screen. The experiment began with four practice trials, and participants were asked if they understood the procedure after the practice trials had been completed. If the gaze contingent trigger failed on any trial, recalibration was undertaken.

### *Data Analysis*

The eye-tracking record for the experimental items was visually inspected using software developed at the University of Massachusetts at Amherst, called EyeDoctor (see footnote 1). Vertical drift was manually corrected to ensure accurate interpretation of the data. Extremely long fixations (greater than 1200 ms) were considered anomalies and removed. Rayner & Pollatsek (1989) showed that readers take up minimal information from extremely short fixations (less than 80ms), so any fixations of 80ms or less were automatically removed, or if within one character of another fixation, merged with that fixation.

---

<sup>1</sup> Downloadable from <http://www.psych.umass.edu/eyelab/software/>.



The experimental items were broken down into six regions. The first region comprised the context sentences (Context region). The third (non-context) sentence was broken down into five regions. The first included the initial noun phrase (NP1). The second was the verb (V). The third was the disambiguating word “by” (Critical region). The fourth was the second noun phrase (NP2). The fifth region was the ten or so characters following the critical region, rounded to the nearest word (Spillover region). An example of the region breakdown is given in 3.2.e:

### 3.2.e. Region Breakdown

/The Smith twins, still only babies, are no... frightened one of the babies but not the other. /

/ 1 /

/The baby /who was /frightened /by /the flames/ was crawling / ...

/ 2 / - / 3 / 4 / 5 / 6 /

1 = Context, 2 = NP1, 3 = Verb, 4 = “by”, 5 = NP2, 6 = Spillover

With the disambiguating word “by” being a function word, and the region it is associated with so small, we were concerned that it would be processed para-foveally from the verb region and not fixated on directly. In order to avoid missing the fixations in which it was processed an additional region was constructed by combining 3 and 4 (the verb and the disambiguating word) together. This region will be referred to as *Verb+By*. The region containing the relativizer ‘who was’ was not analysed as it only appeared in two of the four conditions.

We report data for five fixation duration measures: *First Pass*, *Second Pass*, *Total Time*, *Regression Path*, and *First Fixation*; and three eye movement control measures: *Fixation Probability*, *Regressions In*, and *Regressions Out*. Details of our definitions of each duration and control measure can be found in 3.2.f and 3.2.g. Two sets of software were used to extract the measures above. Primarily the Sturt Suite (developed by Dr. Patrick Sturt, University of Edinburgh) was used, but this was supplemented by EyeDry (developed at the University of Massachusetts at Amherst, see footnote 1).

### 3.2.f. Reading Measures

*First Pass (FP)*: The sum of all fixations made in a region until the first saccade (rightward movement) or regression (leftward movement) exits the region.

*Second Pass (SP)*: The sum of all fixations made in a region after the first time it has been exited left or right.

*Total Time (TT)*: The sum of all fixations made in a region (effectively the sum of First Pass and Second Pass)

*Regression Path (RP)*: The sum of all fixations made once a region has been entered from the left, until it is exited right. This may include fixations in previous regions that are made after the region has been fixated (regressions).

*First Fixation (FF)*: The duration of the first fixation made in a region when it is entered from the left.

### 3.2.g. Fixation Measures

*Fixation Probability (Pfix)*: The proportion of trials in which the region was fixated at least once before the subsequent regions were fixated. The converse of this measure ( $1 - P_{fix}$ ) is Skipping Rate.

*Regressions In (RI)*: The proportion of trials in which the region was fixated after a subsequent region had already been fixated.

*Regressions Out (RO)*: The proportion of trials in which the region was regressed out of before subsequent regions were fixated.

Data for *First Pass* and *Regression Path*, was not included if the region was skipped prior to being fixated but was subsequently fixated in a regression from a region downstream.

Linear Mixed Effects modelling was used to analyse the data. The advantages of LME, as opposed to the traditional ANOVA approach, are well discussed and documented (e.g. Baayen, 2008) but to summarise the most salient points for this circumstance: LME allows for simultaneous consideration of subject and item variability through the use of random effects (Baayen, Davidson, & Bates, 2008; Jaeger, 2008) (essentially replacing the  $F_1$  and  $F_2$  or MinF' analyses). In addition, LME copes with missing data more effectively than ANOVA, and so is particularly suited to eye-tracking research where missing data is a common occurrence.

First we built a baseline model which included the fixed effects and their interactions together with random intercepts by participants and by items. To this model we iteratively added random slopes, corresponding to each of the fixed effects, first by subjects then by items. If a model with random slopes for fixed effects better fit the data then it became the new baseline model for further likelihood ratio tests with random slopes for interactions.

Additionally, a second model was created by applying a log transform to the data during model creation. Note that it is possible to add subject and item random slopes to a log transformed model but that doing so does not appear, at least with our data, to improve the model fit. This may be because the log transformation process reduces the impact of subject and item variability such that random slopes (which represent subject and item variability) no longer capture any variance.

For all duration measures (3.2.f) except *Second Pass*, if there was no data for a region (i.e. the region had been skipped), the zero value was treated as an NA so that the model estimates weren't biased by missing observations. In *Second Pass* it is theoretically interesting to include information about whether the region was skipped (because it means that the region was never re-fixated after the first pass fixation) and so zeroes were not removed.

Scatter plots with best fit lines were observed for the residuals of both best fitting normal model and the log transformed model to identify which was superior. Log transformation is only possible with continuous data that does not contain zeroes, so is only applicable to the *First Pass*, *Total Time*, *Regression Path*, and *First fixation* measures. In these cases the log

transformed model always proved superior to the normal model, and so any data reported for these measures is log transformed. Any models reported for *Second Pass* are not log transformed (due to zeroes being present in the data) and so may include random slopes. The same is true for the fixation measures, all of which provide categorical data (Probability of Fixation, Regressions Out, and Regressions In). For these measures we used binomial linear mixed effects modelling.

Factor levels were converted to numeric values and centred on 0 in order to reduce collinearity between variables (Baayen, 2008). In the case of the LME results, T values were considered to be worth checking for marginality ( $p > 0.05$  but  $< 0.1$ ) if they exceeded 1.85, and for significance ( $p < 0.05$ ) if they exceeded 2 (based on Baayen's estimate of  $t \geq 2$  (Baayen, 2008)). Significance was checked with Markov Chain Monte Carlo sampling if random slopes were not included in the model, and the equation given in 3.2.h (see Baayen, 2008, p269/270) was applied if they were (Markov Chain Monte Carlo sampling is not currently compatible with random slopes due to uncertainty concerning how to calculate degrees of freedom).

### 3.2.h. Formula for checking significance of models with slopes

$$2 * (1 - pt(abs(?tval), ?obs - ?df))$$

?tval = the t value

?obs = total number of observations

?df = degrees of freedom (total number of effects in model: fixed effects, interactions, and random slopes)

### 3.3 Results

**/The baby /who was /frightened /by /the flames/ was crawling / ...**

**/   2   /   -   /   3   / 4 /   5   /   6   /**

**2 = NP1, 3 = Verb, 4 = “by”, 5 = NP2, 6 = Spillover**

#### **Overview of Critical Findings:**

Context and ambiguity interact in the proportion of regressions out of the NP2 region. The interaction is such that no difference in the proportion of regressions out is observed when the relativizer is present (unambiguous) regardless of context. When the relativizer is absent (ambiguous) however, a supportive context reduces the impact of the garden path (as indexed by a lower proportion of regressions out). See Figure 3.1 for details. We also observe a similar Regressions In interaction on the ‘by’ region. See Figure 3.2 for details.

#### **Context by Ambiguity Interactions:**

##### **Regressions Out of the NP2 Region:**

We observe an interaction between context and ambiguity in the NP2 region ( $p = 0.023^*$ ) in which more regressions out occur if the structure is ambiguous compared to unambiguous (see Figure 3.1). Critically the magnitude of this effect varies dependant on the context. If the context is supportive we see a numerical (but not significant) difference (Ambiguous = 22% vs. Unambiguous = 17%,  $p = 0.0769^m$ ), but if the context is not supportive we see a much larger, highly significant difference (Ambiguous = 35% vs. Unambiguous = 15%,  $p < 0.001^{***}$ ). So, unsurprisingly, participants appear to encounter the most difficulty with the ambiguity when the correct relative clause reading is not supported by the context. This implies that the availability of a supportive context helps to reduce the impact of the garden path effect caused by the ambiguity.

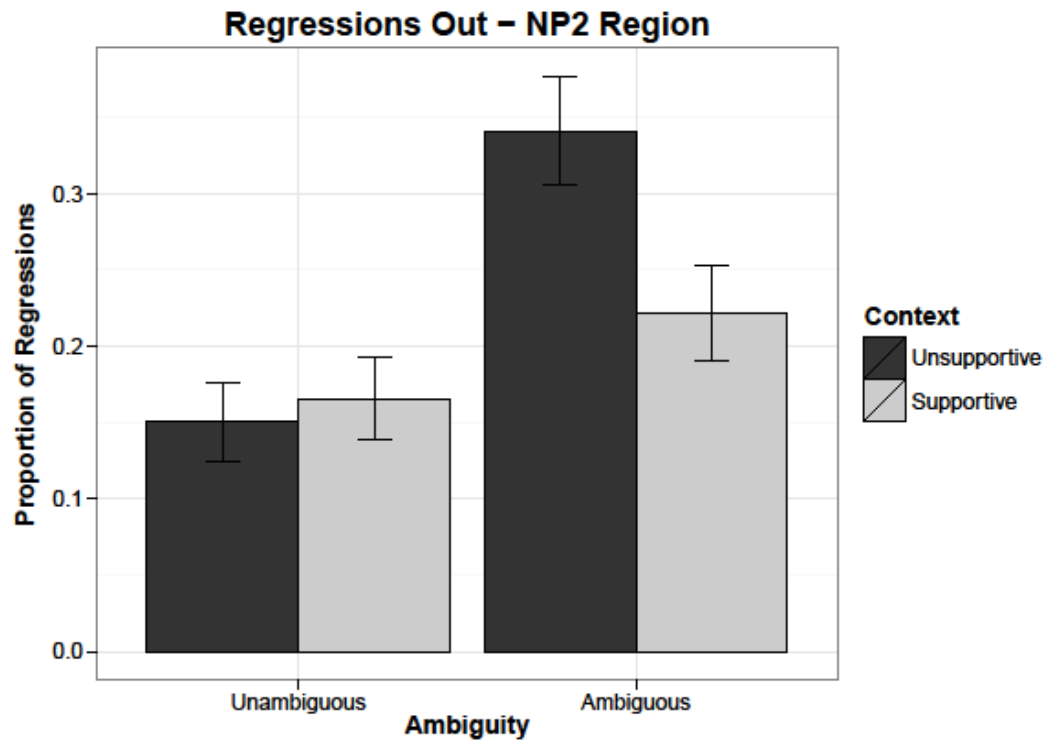


Figure 3.1: Experiment 1- Proportion of Regressions for Regressions Out on the NP2 Region. Error Bars display Standard Error. Context x Ambiguity Interaction.

We observe a non-interactive main effect of ambiguity in the preceding regions (Verb and Verb+By) which indicates that participants are garden-pathed regardless of context. We then observe the above interaction between context and ambiguity which shows that people are more severely garden pathed if the context does not support the correct interpretation of the ambiguity. Taken together these findings appear to support the syntax-first model of parsing as people are initially influenced by the ambiguity, and then are able to utilise the context later on to facilitate a repair (cf. Mitchell, Corley, & Garnham, 1992). Some parallel models, however, can also explain this effect as a cost associated with switching between maintained interpretations. In this thesis we remain agnostic about this issue, as it does not impact the central drive of the research.

#### Regressions In to the 'by' Region:

In Regressions In we observe an interaction between context and ambiguity in the region containing the disambiguating word 'by' ( $p = 0.026 *$ ). The pattern is almost identical to

that observed for the regressions out interaction detailed above. When the relativizer is present we see more regressions in to the disambiguating ‘by’ region when the context is supportive than when it is not (13% vs. 11%), but when the relativizer is not present we see the reverse (Supportive = 16% vs. Unsupportive = 26%). The similarity of the patterns is striking, implying circumstantially at least that the majority of the regressions out from the NP2 region are landing in the disambiguating ‘by’ region. The raw numbers of regressions, seen in the table below (Table 3.1) support this tentative circumstantial assertion.

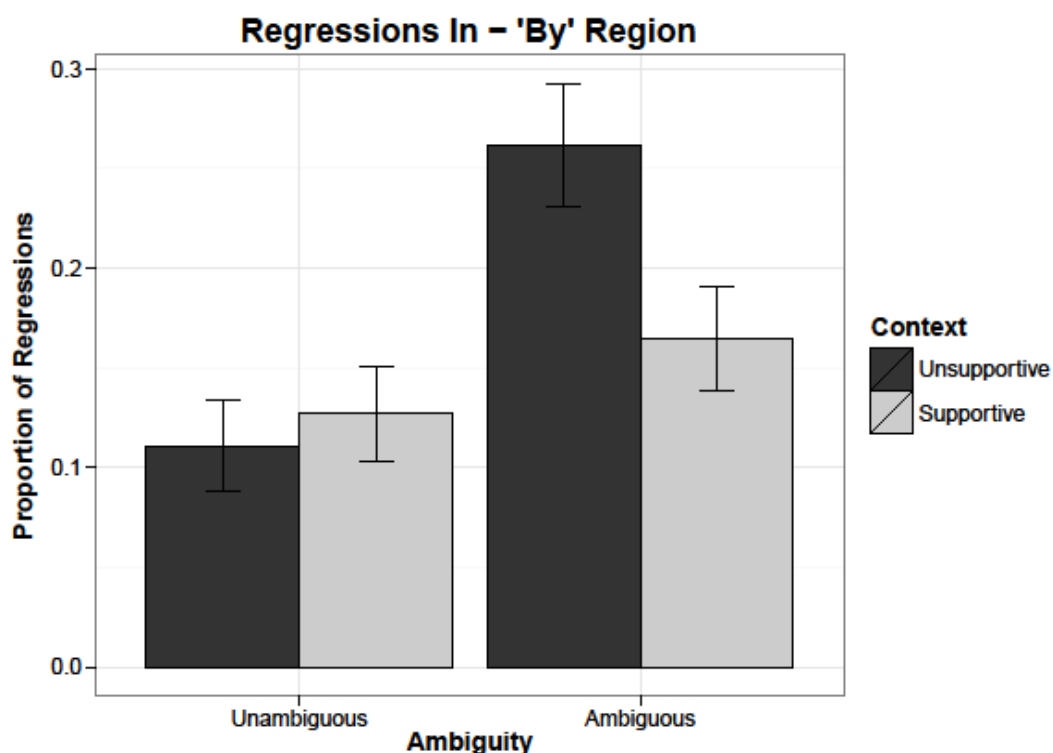


Figure 3.2: Experiment 1 - Proportion of Regressions for Regressions In on the ‘by’ Region. Error Bars display Standard Error. Context x Ambiguity Interaction.

Table 3.1: Experiment 1 - Raw numbers of regressions out of the NP2 region, and raw numbers of regressions in to the ‘by’ region. By condition.

Context by Ambiguity	Regressions Out of NP2		Regressions In to ‘by’	
	Unambiguous	Ambiguous	Unambiguous	Ambiguous
Unsupportive	27	63	32	75
Supportive	31	40	36	48

### Summary of Main Effects:

The main effects are reported in the tables below. Table 3.2 reports Context effects, while Table 3.3 reports Ambiguity effects.

#### Context:

As can be seen from Table 3.2, a supportive context consistently reduces fixation durations on the Verb and the Verb+By regions in early reading measures (FP, FF, RP). This is not so apparent in Second Pass reading times, implying that the influence of a supportive context occurs primarily during initial readings. Perhaps unsurprisingly, we observe a greater proportion of regressions in to the NP1 and Verb+By regions when the context is Unsupportive, compared to when it supports the relative clause reading. In both possible interpretations of the syntax (relative clause or main clause) a supportive context will facilitate processing: In the relative clause case by increasing discourse cohesion, and in the main clause case, by aiding in reanalysis and recovery from the garden path.

When the context is Unsupportive, we also see a higher probability of fixating the spillover region and a marginally higher probability of fixating the Verb, but it is worth being cautious about both of these effects as one is marginal and the other is very close to ceiling.

#### Ambiguity:

As can be seen from Table 3.3, we observe the established garden path pattern: A consistently negative effect on reading times on and around the Verb and disambiguating regions, and increases regressions out of and in to almost all regions. The one exception is a reduction in reading times associated with the First Pass of the Spillover region, when the syntax of the previous regions was ambiguous. This is likely either a function of spending a greater amount of time on the previous regions when they were ambiguous (and thus gaining more preview of the spillover region), or due to those who are garden pathed being more likely to regress out of that region to return to the source of difficulty. It may also be considered a classic spillover effect (Just, Carpenter, & Woolley, 1982).



Table 3.2: Experiment 1 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if Context is **Unsupportive** or **Supportive**. Effects are also marked if **Marginal**.

<b>Context</b>	<b>Levels</b>	<b>NP1</b>		<b>Verb</b>		<b>By</b>		<b>NP2</b>		<b>Spillover</b>		<b>Verb+By</b>		<b>By+NP2</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unsupportive	0.35	336.7	-3.08	277.1	-1.43	252.1	-0.64	293.3	1.23	904.5	-2.39	299.21	-0.97	341.9
	Supportive		373.4	**	250.3		225.2		287.1		940.5	*	271.67		329.5
First Fixation	Unsupportive	0.73	233.9	-2.29	236.2	-1.50	244.0	-0.10	221.3	-0.87	240.9	-2.68	235.61	-0.31	226.1
	Supportive		226.1	*	222.4		215.8		219.8		237.3	**	220.03		222.8
Regression Path	Unsupportive	0.48	427.1	-3.49	479.3	0.14	353.5	-1.73	419.0	-0.79	1134	-2.70	517.33	-1.07	467.4
	Supportive		439.7	***	371.8		449.1		426.2		1256	**	434.07		497.5
Second Pass	Unsupportive	0.55	160.7	0.20	153.9	-1.00	44.73	0.42	105.8	-0.57	233.0	-1.94	207.37	0.18	136.8
	Supportive		186.5		156.8		38.45		111.4		214.9	<sub>m</sub>	198.66		139.9
Total Time	Unsupportive	0.30	432.4	-3.07	419.1	0.11	257.6	-0.22	393.6	0.00	1128	-3.38	493.90	-1.00	470.2
	Supportive		495.5	***	384.8		260.9		390.4		1153	***	451.07		457.9
Probability of Fixation	Unsupportive	1.04	.72	-1.84	.83	-0.19	.18	-0.39	.90	-2.25	.99	-1.56	.89	-1.12	.94
	Supportive		.75	<sub>m</sub>	.79		.18		.89	*	.96		.86		.93
Regressions Out	Unsupportive	-1.36	.07	-0.42	.24	0.54	.20	-1.53	.25	-0.14	.29	-0.28	.24	-0.70	.19
	Supportive		.04		.23		.23		.19		.29		.23		.17
Regressions In	Unsupportive	-2.13	.45	-0.74	.35	-1.33	.19	0.73	.26	0.60	.17	-2.29	.44	0.38	.29
	Supportive	*	.39		.33		.15		.28		.18	*	.37		.30

Table 3.3: Experiment 1 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if the local Syntax is Ambiguous or Unambiguous. Effects are also marked if Marginal.

<b>Ambiguity</b>	<b>Levels</b>	<b>NP1</b>		<b>Verb</b>		<b>By</b>		<b>NP2</b>		<b>Spillover</b>		<b>Verb+By</b>		<b>By+NP2</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unambiguous	-1.32	371.6	3.25	247.3	0.95	230.2	0.60	286.9	-2.12	939.5	3.68	263.1	2.16	323.4
	Ambiguous		338.5	***	280.0		247.2		293.5	*	905.6	***	307.7	*	348.0
First Fixation	Unambiguous	-1.00	234.1	1.73	222.1	0.72	223.1	-0.01	219.2	1.30	235.0	1.68	220.9	0.49	221.8
	Ambiguous		225.8		236.5		236.7		221.9		243.2		234.6		227.1
Regression Path	Unambiguous	0.13	416.6	4.79	396.1	1.55	347.2	4.96	390.9	2.23	1238	5.31	427.8	5.56	434.5
	Ambiguous		450.1	***	455.0		455.3	***	454.3	*	1353	***	523.5	***	530.3
Second Pass	Unambiguous	2.83	148.6	6.64	95.96	3.68	26.87	2.93	86.49	2.00	192.7	4.82	128.9	3.47	104.3
	Ambiguous	**	198.5	***	214.8	***	56.32	**	130.7	*	255.3	***	277.1	***	172.4
Total Time	Unambiguous	2.44	455.4	9.90	332.6	3.74	232.8	4.10	367.9	1.19	1128	11.42	380.6	6.28	418.8
	Ambiguous	*	472.4	***	471.3	**	285.7	***	416.0		1153	***	564.3	***	509.3
Probability of Fixation	Unambiguous	-1.23	.75	3.38	.77	0.71	.17	0.33	.89	0.11	.97	2.59	.85	-0.74	.94
	Ambiguous		.72	***	.85		.19		.90		.98	**	.90		.93
Regressions Out	Unambiguous	0.66	.05	2.37	.20	1.16	.18	4.58	.16	2.36	.26	2.60	.20	4.14	.13
	Ambiguous		.06	*	.27		.24	***	.28	*	.32	**	.27	***	.23
Regressions In	Unambiguous	2.38	.38	5.47	.26	4.09	.12	2.20	.24	-0.04	.17	6.24	.31	2.41	.27
	Ambiguous	*	.45	***	.42	***	.21	*	.30		.17	***	.50	*	.33

### 3.4 Discussion

In this experiment we have established a baseline for participants reacting to syntactic ambiguity with and without the presence of a supportive context. We observe main effects of ambiguity and context (Verb & Verb+By regions) in early reading measures (FP, FF & RP) such that longer reading times are observed when the context is unsupportive, and when there is an ambiguity present, but critically these variables do not interact. We see also similar effects of ambiguity (all regions) in later reading measures (SP & TT) but no late effects of context. Of particular interest to us going forward are the context and ambiguity effects in the Regression measures: In practically all regions we see ambiguity effects for both regressions in and out, such that there are more regressions made in the ambiguous condition. Context effects are more limited, appearing only in Regressions In to the NP1 and Disambiguating (Verb+By) regions, and showing more regression in when the context is not supportive. The overall pattern emerges that a supportive context facilitates parsing regardless of ambiguity, and that an ambiguity causes difficulty, regardless of context.

Critically, however, we see the two variables interact in the predicted fashion in Regressions Out of the NP2 region (the baby frightened by the flames). The interaction shows that participants are able to utilise a supportive context to facilitate the processing of a syntactic ambiguity.

We remain agnostic as to whether the results better support syntax-first or parallel models of parsing, as the theoretical distinction of whether the observed difficulty / facilitation is due to a repair process or a selection cost is not relevant to this thesis.

Importantly for this thesis, we have established a baseline against which we can compare the reading behaviour of participants performing the experiment while under an extrinsic memory load.

## **Chapter 4: Load, Context, and Syntactic Ambiguity (Experiment 2)**

### **4.1 Introduction**

In the previous experiment we saw strong main effects indicating that processing of a sentence was facilitated by the addition of a prior referential contrast set. We also saw that participants favoured by default the main clause interpretation when encountering reduced relative clauses.

Context and Ambiguity interactions observed in the NP2 region for Regressions Out, and the disambiguating 'by' region for Regressions In, indicated that participants were better able to resolve a syntactic ambiguity if the context provided support for the correct interpretation.

In this experiment we will replicate the procedure of the previous experiment but apply a high titrated extrinsic memory load to participants to observe how people with experimentally reduced cognitive resources process contextual information and apply it to the resolution of syntactic ambiguities. Particularly, we will be interested to see what effect cognitive loading has on the interactions between context and ambiguity which were observed in the previous experiment.

Following the High / Low span reader distinctions of King & Just (1991) and MacDonald et al. (1992) we predict that participants under load will not be able to make use of contextual information in order to assist with the parsing of syntactic ambiguities *online*.

We therefore expect to see main effects of ambiguity as before (increased reading times when ambiguous as compared to when unambiguous) but a reduction / elimination of the observed context effects and ambiguity by context interactions.

## 4.2 Methods

### *Participants*

36 participants took part in the study. The requirements were the same as experiment 1, with the addition that participants had not taken part in that experiment.

As in the NoLoad experiment, an 80% threshold for comprehension question accuracy was imposed. In addition, participants were required to recall 70% of the digits accurately (see below for details) as we had to be sure that attention was being paid to both tasks.

Total Exclusions based on poor comprehension, tracking, or digit recall = 5

Comprehension Question Range (of included participants) = 85% - 98%, Mean = 92%.

Digit Span Range (of included participants) = 3 – 9, Mean = 6.

### *Stimuli*

In all ways, the stimuli were identical to those in experiment 1. The digits for the recall task were integrated into the experiment in a pseudo random fashion, so that even if several participants were set to a difficulty of five digits (for example), no two of those participants would see the same digits preceding a given sentence. This was done to make sure that there was no unintentional structured interference from the digits, on the reading task.

### *Procedure & Data Analysis*

In all ways the procedure was identical to that in experiment 1 with the exception of the digit span pre-test, and subsequently differing trial presentation. The pre-test was designed to establish an appropriate level of cognitive loading for each individual participant and was a simple program, written in Visual Basic for applications (VBA), which emulated the presentation of the main experiment (see Chapter 2 for details). Participants were presented with a series of digits and asked to memorise them. When the participant pressed the 'Next' button the digits were removed from the screen and a sentence appeared. The participants were instructed to read the sentence silently and press 'Next'

when done. Upon pressing 'Next' the sentence disappeared and the participants were asked to recall the digits. The accuracy of the recall was noted by the experimenter, and the participants were then presented with a yes / no comprehension question, the accuracy of their answers was also recorded. The VBA program allowed for the adjustment of the number of digits presented to a participant (3-9) and the decision as to whether to increase, decrease, or maintain the difficulty was made by the experimenter. A participant was required to achieve accuracy on both the digits and the comprehension questions (in the case of the digits, this meant recalling the correct digits in the correct order, in the case of the comprehension question this meant a correct answer). Each level of difficulty (indexed by the number of digits presented) was repeated until that the participant was accurate or inaccurate twice. If accuracy was achieved twice then the difficulty was increased until the participant was no longer reliably accurate. A level was selected when a participant was able to be reliably accurate at that level, but not at the next level up.

The presentation of the main experiment followed the structure of the pre-test. First, digits were presented. Then, as in experiment 1, a gaze contingent trigger was placed in the vertical centre but at the left side of the screen. Upon fixation this was replaced by a filler or experimental item. After the reading component a prompt for the digit recall was presented, and participants were required to say the digits out loud. Finally the comprehension question was presented and participants answered this, as in experiment 1, using the gamepad. A microphone captured the digit recall, and the experimenter checked off the digits against a list, using different symbols for accuracy, inaccuracy due to wrong digits, and inaccuracy due to wrong order. A fourth symbol was used to indicate if the participant was only one digit out, or had transposed two digits. Totals were made after the experiment for each of the symbols, as well as an overall accuracy total in which accurate answers were totalled and the three types of inaccuracy were conflated into one. Participants were instructed, prior to the experiment starting, that they should be focusing equally on each task and that they were required to achieve overall thresholds of correctness on both the comprehension questions and the digit recall, or their data would be discarded. In this way we aimed to avoid issues of biased attention towards the digit task (see Literature Review, section 1.2.4.2).

In order to maintain comparability, the same procedures were used to analyse experiment 2, as were used to analyse experiment 1.

### 4.3 Results

**/The baby /who was /frightened /by /the flames/ was crawling / ...**

**/   2   /   -   /   3   / 4 /   5   /   6   /**

**2 = NP1, 3 = Verb, 4 = “by”, 5 = NP2, 6 = Spillover**

#### **Overview of Critical Findings:**

With this experiment, our intent was to see if the contextual facilitation of ambiguity processing seen in the previous (NoLoad) experiment would be reduced or eliminated by the addition of an extrinsic memory load. This manifested as interactions between Context and Ambiguity in both the proportions of Regressions Out of the NP2 region, and the proportions of Regressions In to the ‘by’ region. Neither interaction was observed in this experiment which implies that readers under Load are less able to make use of contextual information to assist in the processing of Lexical Ambiguities. Supporting this assertion is a general reduction in main effects of context compared to the previous NoLoad experiment. Although it is worth noting that some context effects are still present so any influence of Load is a reduction rather than elimination.

#### **Context by Ambiguity Interactions:**

Probability of Fixating the Spillover Region:

In the Spillover region we observe a small but reliable two way interaction between Ambiguity and Context ( $p = 0.026^*$ ) (See Figure 3.1). When the relativizer is present (unambiguous) we see a greater probability of fixation if the context is unsupportive, compared to if it is supportive (98% vs. 95%). Whereas when the relativizer is absent (ambiguous) we see the opposite pattern: a higher likelihood of fixating the spillover region if the context is supportive, compared to when it is unsupportive (99% vs. 97%). It is worth considering this interaction cautiously as the differences are very small and at ceiling. This means that we have very many observations of the region being fixated but very few of it being skipped. We cannot be confident in the meaningfulness of the few observations of

skipping and so, despite the regression reporting it as significant, we consider this interaction more likely to be just random noise.

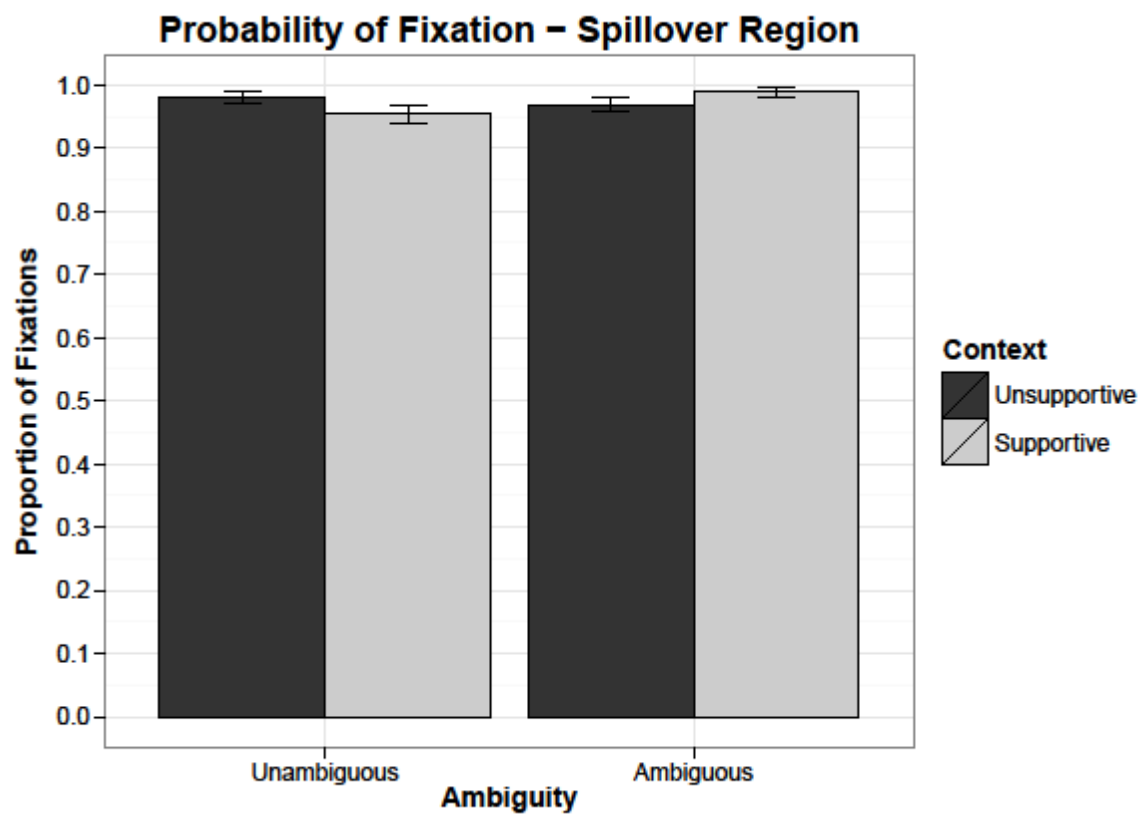


Figure 4.1: Experiment 2 - Proportion of Fixations for Probability of Fixation on the Spillover Region. Error Bars display Standard Error. Context x Ambiguity Interaction.



### Summary of Main Effects:

The main effects are reported in the tables below. Table 4.1 reports Context effects, while Table 4.2 reports Ambiguity effects.

#### Context:

As can be seen in Table 4.1, there are fewer contexts effects observed in this experiment. In the early reading measures we see a regression path for the 'by' region which patterns in the expected direction (facilitation of a supportive context), but a first fixation effect (on the Verb) which patterns the opposite. We also see a probability of fixation effect for this region such that readers are more likely to fixate the Verb if the context is unsupportive. This provides a confusing pattern in which participants are, when the context is unsupportive, more likely to fixate the Verb but will spend less time looking at it (in their first fixation at least). It is possible that the greater attraction to the Verb in the unsupportive condition causes a poorly planned saccade which requires adjusting after landing. We have not conducted a micro-saccade analysis however, so this explanation is purely speculation.

#### Ambiguity:

Main effects of ambiguity remain similar to the previous experiment (see Table 4.2). Notably, the first pass effect in the spillover region which runs counter to the main trend is present here, as it was before. Conversely, we see ambiguity effects in the probability of fixating the NP2 region (and the By+NP2 region) which were not previously observed. In both cases there is a higher likelihood of fixating the region if the preceding syntax has not been ambiguous. This is likely due to participants spending more time in the previous regions if they have been garden pathed, and taking up enough information about the NP2 in preview to feel they can skip it. This explanation is supported by longer first pass reading times in the Verb+By region when in the ambiguous conditions.

Table 4.1: Experiment 2 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if Context is **Unsupportive** or **Supportive**. Effects are also marked if **Marginal**.

<b>Context</b>	<b>Levels</b>	<b>NP1</b>		<b>Verb</b>		<b>By</b>		<b>NP2</b>		<b>Spillover</b>		<b>Verb+By</b>		<b>By+NP2</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unsupportive	0.83	333.1	0.08	253.5	-1.64	230.4	-0.74	266.0	-0.52	787.5	0.06	270.0	0.05	298.4
	Supportive		385.4		250.2		201.2		268.3		793.1		263.4		304.1
First Fixation	Unsupportive	-1.83	241.7	2.15	217.8	-1.52	226.3	-0.34	220.9	0.20	228.8	1.56	218.9	-0.57	221.8
	Supportive		229.2	*	234.2		213.5		128.7		230.7		228.9		218.6
Regression Path	Unsupportive	1.06	411.3	0.06	368.1	-2.19	420.2	-1.49	388.0	-0.84	1232	-0.55	410.8	-1.29	443.9
	Supportive		526.4		378.5	*	292.4		397.4		1234		400.2		439.2
Second Pass	Unsupportive	0.62	192.3	0.59	168.9	-0.49	44.15	-1.09	164.4	-1.80	232.1	0.48	214.2	-1.28	203.6
	Supportive		230.4		181.2		40.16		145.9		258.8		223.7		179.7
Total Time	Unsupportive	0.87	464.1	-0.27	411.3	-0.97	261.4	-1.16	426.2	-1.94	1103	-0.39	469.4	-1.16	496.9
	Supportive		567.3		427.7		249.3		409.0	<sup>m</sup>	1056		479.9		475.3
Probability of Fixation	Unsupportive	1.96	.68	-2.83	.78	1.64	.15	-0.14	.84	0.11	.97	-2.13	.85	0.28	.88
	Supportive	<sup>m</sup>	.72	**	.70		.18		.83		.97	*	.80		.88
Regressions Out	Unsupportive	0.06	.05	0.12	.18	-0.54	.18	-0.66	.19	-1.27	.33	-0.50	.19	-0.11	.16
	Supportive		.06		.18		.14		.17		.30		.17		.15
Regressions In	Unsupportive	-0.34	.47	0.30	.38	-0.56	.17	-0.56	.34	-0.59	.24	-0.41	.44	-0.36	.36
	Supportive		.40		.38		.15		.32		.23		.42		.36

Table 4.2: Experiment 2 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if the local Syntax is Ambiguous or Unambiguous. Effects are also marked if Marginal.

<b>Ambiguity</b>	<b>Levels</b>	<b>NP1</b>		<b>Verb</b>		<b>By</b>		<b>NP2</b>		<b>Spillover</b>		<b>Verb+By</b>		<b>By+NP2</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unambiguous	0.95	361.5	2.80	238.0	0.73	213.5	-0.07	268.6	-3.20	828.9	3.43	250.4	0.48	301.0
	Ambiguous		357.0	**	265.6		218.1		265.7		**	751.7	***	283.0	301.5
First Fixation	Unambiguous	0.55	234.0	1.26	220.3	0.47	213.5	-0.07	219.6	-0.47	231.4	1.23	219.3	0.27	219.4
	Ambiguous		236.9		231.7		213.2		220.0		228.1		228.5		221.0
Regression Path	Unambiguous	0.32	499.8	5.68	327.2	1.90	299.0	2.50	391.7	2.70	1179	6.62	343.0	3.67	424.7
	Ambiguous		437.9	***	419.5	m	413.6	*	393.7	**	1287	***	468.0	***	458.4
Second Pass	Unambiguous	2.71	180.9	5.76	117.2	3.67	29.31	3.18	129.0	2.68	242.8	7.43	145.7	3.93	149.3
	Ambiguous	**	241.7	***	232.9	***	55.44	**	181.3	**	339.1	***	292.1	***	234.0
Total Time	Unambiguous	3.01	493.4	6.27	357.5	1.94	240.7	3.45	396.6	0.85	1075	8.62	390.7	5.07	449.0
	Ambiguous	***	538.1	***	481.6	m	270.0	***	438.6		1085	***	558.6	***	523.2
Probability of Fixation	Unambiguous	-1.27	.71	0.44	.74	1.18	.15	-1.96	.86	1.06	.97	-0.44	.83	-2.37	.90
	Ambiguous		.68		.74		.18	*	.82		.98		.82	*	.86
Regressions Out	Unambiguous	-0.47	.06	4.58	.12	2.45	.09	3.33	.13	4.12	.26	5.33	.11	4.13	.11
	Ambiguous		.05	***	.25	*	.23	***	.22	***	.37	***	.25	***	.20
Regressions In	Unambiguous	3.75	.38	6.97	.28	3.57	.12	3.93	.28	0.26	.23	6.52	.34	4.24	.30
	Ambiguous	***	.49	***	.48	***	.20	***	.38		.24	***	.52	***	.42

## 4.4 Discussion

The interpretation of this experiment is very straightforward. We observe the anticipated ambiguity effects (longer reading times when there is an ambiguity present than when not) implying that regardless of the memory load that they are under, participants are still susceptible to the garden path effects of a syntactic ambiguity, and do attempt to process it correctly / revise their incorrect interpretation (unlike the low span readers in the MacDonald et al., 1992, experiments). Participants are, however, unable to make use of the referential contrast sets provided to assist in the processing of the ambiguity. The only interaction between context and ambiguity that we see is most likely a spurious ceiling effect in Probability of Fixation in the spillover region. Of particular note is that we do not see context and ambiguity interactions in the Regressions Out of the NP2 region, or the Regressions In to the disambiguating 'by' region, like the ones observed when participants were not loaded.

Although there are many fewer main effects of context in this experiment, it does seem that some contextual information is being utilised to facilitate parsing (if not ambiguity resolution) as, with the exception of one, those main effects that we do see pattern as before (shorter reading times when the context is supportive).

These results support our predictions that participants under cognitive load are not able to make use of contextual information to facilitate syntactic ambiguity resolution. However, it is of interest that a). participants were still affected by ambiguities (showing they do not act quite the same as the low span readers of MacDonald et al., 1992), and b). contextual information was still made use of sporadically to facilitate the parsing of non-ambiguous text. This implies that while contextual information was available and sometimes made use of in general parsing, participants under load did not apply it in the resolution of syntactic ambiguities.

The fact that low span readers and loaded readers do not act the same lends credence to the possibility that higher span readers have more efficient mechanisms, rather than a greater capacity (see Literature Review, section 1.3). If it was simply a capacity issue then the reduction in the availability of resources should have the effect of reducing a reader's capacity to that of low span reader. However, if you reduce the available capacity of a

reader that does not necessarily mean that you remove the skill of contextual processing which may be more highly developed as a result of practice. They may be able to still utilise that skill (albeit at a reduced level) even under the influence of load.

Overall, however, this experiment supports the Capacity Constrained model of syntactic processing: A reduction in resources reduces the reader's ability to apply non-syntactic information to aid in parsing. This is particularly evident when parsing becomes more difficult due to the presence of a syntactic ambiguity.

## Chapter 5: Combined Analysis of NoLoad and Load Experiments

### 5.1 Introduction

In the previous sections we have made assertions as to the differences between the two experiments above and allowed them to guide our understanding of the contextual facilitation of syntactic ambiguity resolution in the context of resource availability. Before continuing we will briefly summarise the findings which led us to those assertions:

The tables below show the Context (Table 5.1) and Ambiguity (Table 5.2) main effects that differ between the NoLoad and Load experiments.

Overall, we can see from Table 5.1 that there is general reduction in the use of context of the Loaded participants compared to the non-Loaded participants. This is principally evident on the Verb and Verb+By regions.

Table 5.1: Context Effects which were present / absent between experiments. NL = present in NoLoad experiment, L = present in Load experiment. x = not present.

<i><b>The baby frightened by the flames was crawling away as quickly as she could.</b></i>							
	NP1	Verb	By	NP2	Spillover	Verb+By	By+NP2
First Pass		NL x				NL x	
First Fixation						NL x	
Regression Path		NL x	x L			NL x	
Second Pass							
Total Time		NL x				NL x	
Probability of Fixation					NL x	x L	
Regressions Out							
Regressions In	NL x					NL x	

We do not see the same general trend for ambiguity. When an ambiguity is present we have previously observed longer reading times, a higher probability of fixation, and more regressions in and out. Table 5.2 shows that there is very little general reduction of the negative reaction in the presence an ambiguity for loaded participants as compared to non-loaded participants.

Table 5.2: Ambiguity effects which were present / absent between experiments. NL = present in NoLoad experiment, L = present in Load experiment. x = not present.

<i><b>The baby frightened by the flames was crawling away as quickly as she could.</b></i>							
	NP1	Verb	By	NP2	Spillover	Verb+By	By+NP2
First Pass							NL x
First Fixation							
Regression Path							
Second Pass							
Total Time							
Probability of Fixation		NL x		x L			x NL
Regressions Out			x L				
Regressions In							

Additionally, recall that in the NoLoad experiment we observed a two way Context by Ambiguity interaction in the proportions of Regressions Out of the NP2 region (see Figure 3.1 in section 3.3). The interaction ( $p = 0.023 *$ ) showed participants making use of the available context to facilitate processing of the ambiguity and was yoked with a Regressions In interaction ( $p = 0.026 *$ ). Neither the regressions Out nor the Regression In interactions were present in the Load experiment (see Figure 5.1 and Figure 5.2, below). The pattern for Regressions Out is similar to the one observed for the No Load experiment, but there is a general reduction in the proportions of Regressions Out, and the interaction between Context and Ambiguity is not significant ( $p = 0.118$ ). For the Regressions In interaction, the pattern of the Unambiguous conditions switches such that there are numerically more Regressions In when the context is unsupportive, than when it is supportive. However, the difference between these two conditions in the NoLoad experiment was not significant, and nor is it here ( $p = 0.307$ ). We see the same overall reduction in the proportion of regressions in, as we do for the Regressions Out data above, and the Context by Ambiguity interaction is not significant ( $p = 0.773$ ).

It was the lack of these interactions, combined with the reduction in the use of context (mentioned above) that led us conclude that reducing the resources available to a participant reduced their ability to use context to parse sentences and, particularly, rendered them unable to use context in the resolution of syntactic ambiguities.

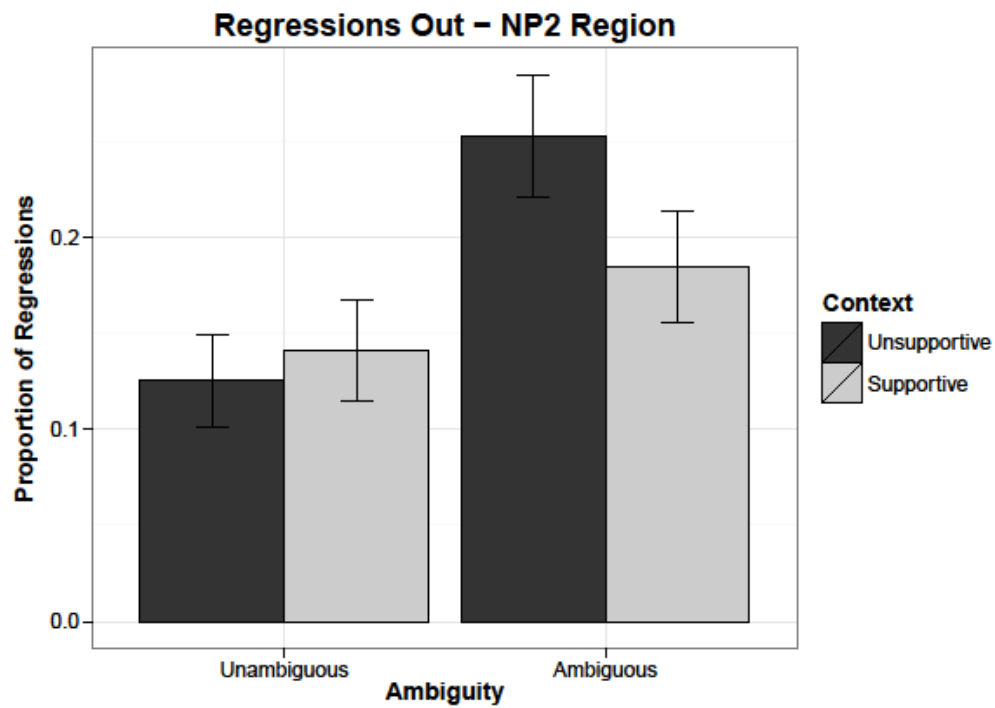


Figure 5.1: Experiment 2 - Proportion of Regressions for Regressions Out of the NP2 Region for the Load experiment. Error bars display standard error. Context x Ambiguity Interaction.

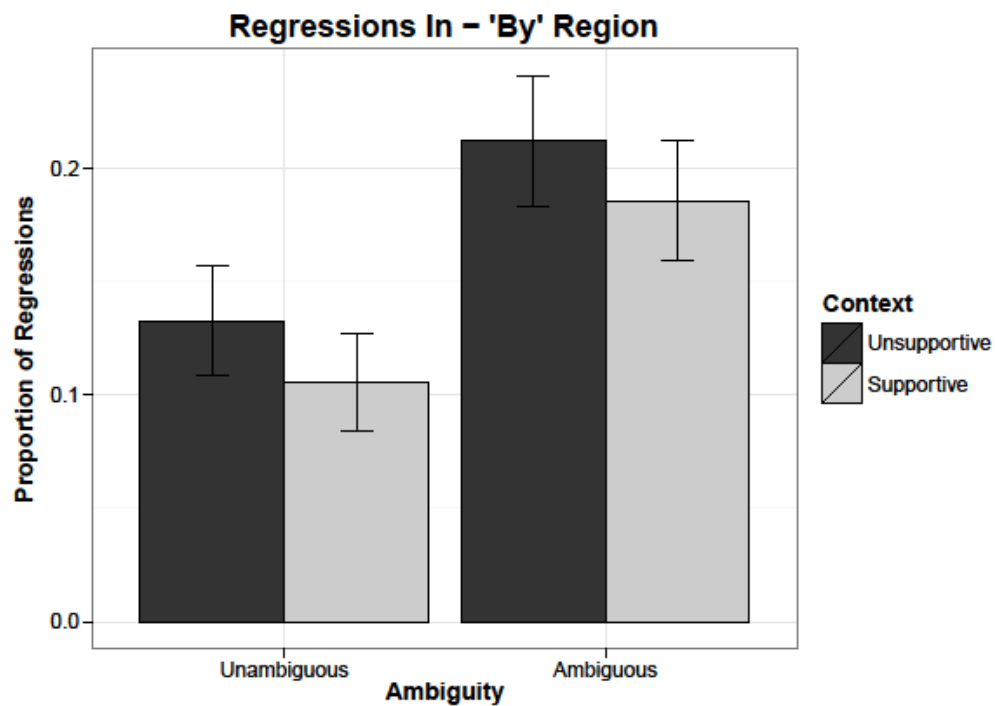


Figure 5.2: Experiment 2 - Proportion of Regressions for Regressions In to the 'by' Region for the Load experiment. Error bars display standard error. Context x Ambiguity Interaction.



Importantly, however, we have not conducted a statistical analysis to see whether the differences we note are reliable.

The data presented in the upcoming sections represents a combination of the data of each experiment, with experiment (Load) being added as a factor into the statistical analysis. If we observe interactions with Load we can conclude that the differences that we observe between the two experiments are real, and not a result of noise in the data.

It is worth noting that the presence of a main effect of Load is not particularly theoretically interesting as it only indicates a difference between the two experimental groups. This difference isn't wholly attributable to our manipulation as the individuals in the second group might have differed from those in the first independently of our attempts to reduce their available resources. Interactions with Load, on the other hand, will be very interesting as they indicate that the two groups differ in a non-general way. Systematic variance which is anticipated by our theoretical approach is extremely unlikely to be a simple artefact of the between subjects design.

Nevertheless, we assumed that the biggest difference between our two experimental groups is likely to be our Load manipulation, so, with caution, we developed a post-hoc measure of between group variance to check that they did differ significantly (see below for details).

## 5.2 Methods

### *Data Analysis*

A two level predictor called Load was created to distinguish between Loading states (and thus experiments). Experiment 1 was indexed with 'NoLoad', while experiment 2 was indexed with 'Load'. The two datasets were then merged together. Analysis then proceeded as before, with the addition of the Load predictor, which was centred as with the other predictors. It is worth noting that by subjects random slopes cannot be applied to the Load factor, and so only by items slopes were used on this predictor. In addition to the addition of the Load factor, analysis of a further fixation measure was conducted to independently test whether the Loading manipulation was affecting reading behaviour. The

analysis examined the total number of fixations made across the critical regions before the most rightwards of them was exited right. For this combined analysis, the critical regions are: pre-critical, critical, and spillover.

### 5.3 Results

**/The baby /who was /frightened /by /the flames/ was crawling / ...**

**/   2   /   -   /   3   / 4 /   5   /   6   /**

**2 = NP1, 3 = Verb, 4 = "by", 5 = NP2, 6 = Spillover**

#### **Number of Fixations:**

We measured the raw number of fixations made in and around the Critical Region as a crude but effective independent measure of the efficacy of the Cognitive Loading Manipulation. We found that there was a main effect of Load ( $p = 0.008^{**}$ ), in which the average number of fixations around the critical region was higher when there was no extrinsic memory load, and lower when there was (2.41 vs 2.14). We view this as evidence that the digit span load manipulation is having an influence on eye movements and reading behaviour, but do not try to reason theoretically about what it might mean beyond that as the measure is post-hoc and rather crude.

Despite this lack of theoretical underpinning we are somewhat confident that this measure represents the effects of Loading (rather than just general differences between experimental groups) because we use it in this between subjects analysis and the one which we report later on involving Lexical Ambiguity. A standard interpretation would be to consider it as an uninformative group difference; however, since a similar pattern is found for the other experiments involving load we suggest, cautiously, that it provides prima facie evidence that the load manipulation affects some aspects of reading behaviour.

### Overview of Critical Findings:

This meta-analysis was conducted to statistically assess the presence of the Regressions Out and Regressions In interactions showing contextual facilitation of ambiguity resolution in the NoLoad experiment, and their subsequent absence from the Load experiment. Statistically we find that the differences between experiments are not reliable, with the Regressions Out interaction being non-significant ( $p > 0.1$ ) and the Regressions In interaction being marginal ( $p = 0.081^m$ ). See Figure 5.4 and Figure 5.5 for details.

### Context by Ambiguity Interactions:

Regressions Out of the NP2 region:

In the NP2 region we observe a two way interaction between Context and Ambiguity ( $p = 0.006^{**}$ ) for Regressions Out (see Figure 5.3).

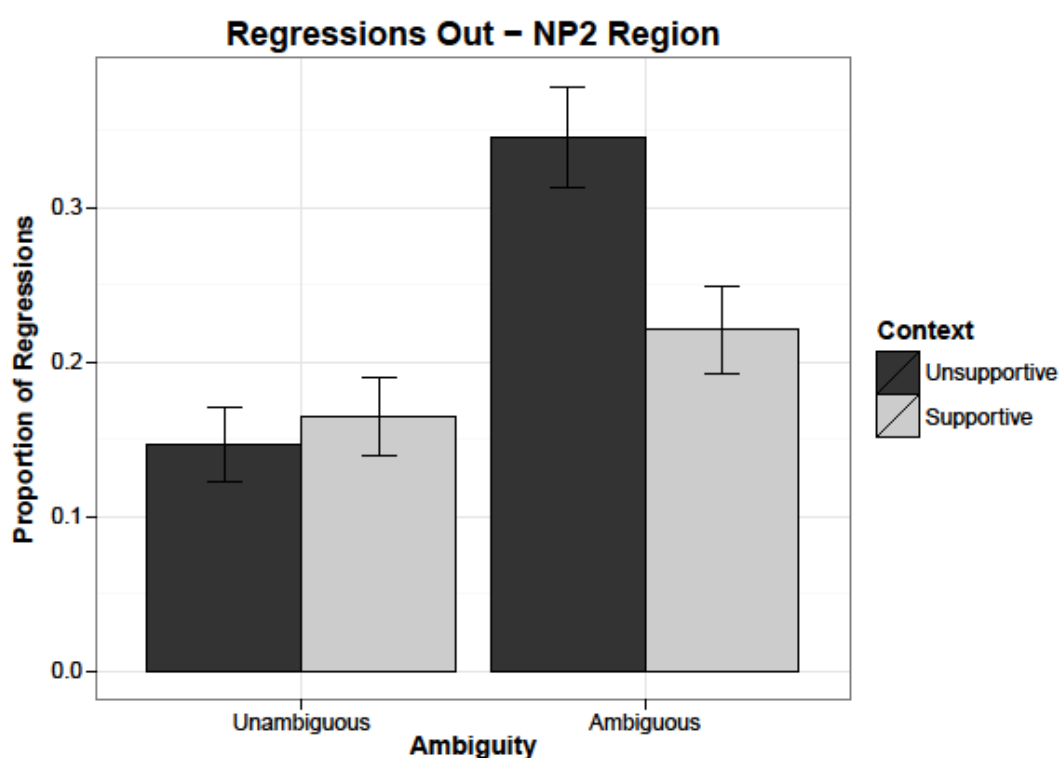


Figure 5.3: Combined Analysis of Experiments 1 & 2. Proportion of Regressions for Regressions Out on the NP2 Region. Error Bars display Standard Error. Context x Ambiguity Interaction.

This is the same pattern as we observed in the NoLoad experiment described above (see Figure 3.1, section 3.3) and the same explanation applies: People are more severely garden pathed if the context does not support the correct interpretation of the ambiguity, and so regress out in the hope that a second reading will resolve the ambiguity.

As we have discussed in the introduction to this chapter this interaction was present in the NoLoad experiment but absent in the Load experiment. We cautiously took this to imply that participants who were not under cognitive load were able to make use of context to facilitate the processing of the ambiguity, while those who were Loaded were not (see section 4.4). However, Load does not interact with this interaction ( $p = 0.732$ , see Figure 5.4) so we cannot claim with any surety that this is the case.

#### Regressions In to the 'by' region:

In the introduction to this chapter we also observed that the Regressions In Context by Ambiguity interaction which, in the No Load experiment, appeared yoked to the Regressions Out Context by Ambiguity interaction, was not present in the Load experiment. Figure 5.5 shows the difference between experiments. The three way Context by Ambiguity by Load interaction was marginal ( $p = 0.081^m$ ). As noted before, the pattern is numerically different for the Unambiguous condition between experiments, but as neither pairwise comparisons (Unambiguous: Unsupportive vs Supportive) were significant (NoLoad, 11% vs. 12%,  $p > 0.1$ , Load, 13% vs. 10%,  $p > 0.1$ ), this is of little interest. Critically, when there is an ambiguity, participants under who are under Load appear less likely to regress in to the 'by' region dependant on the availability of context, than are those who are not Loaded.

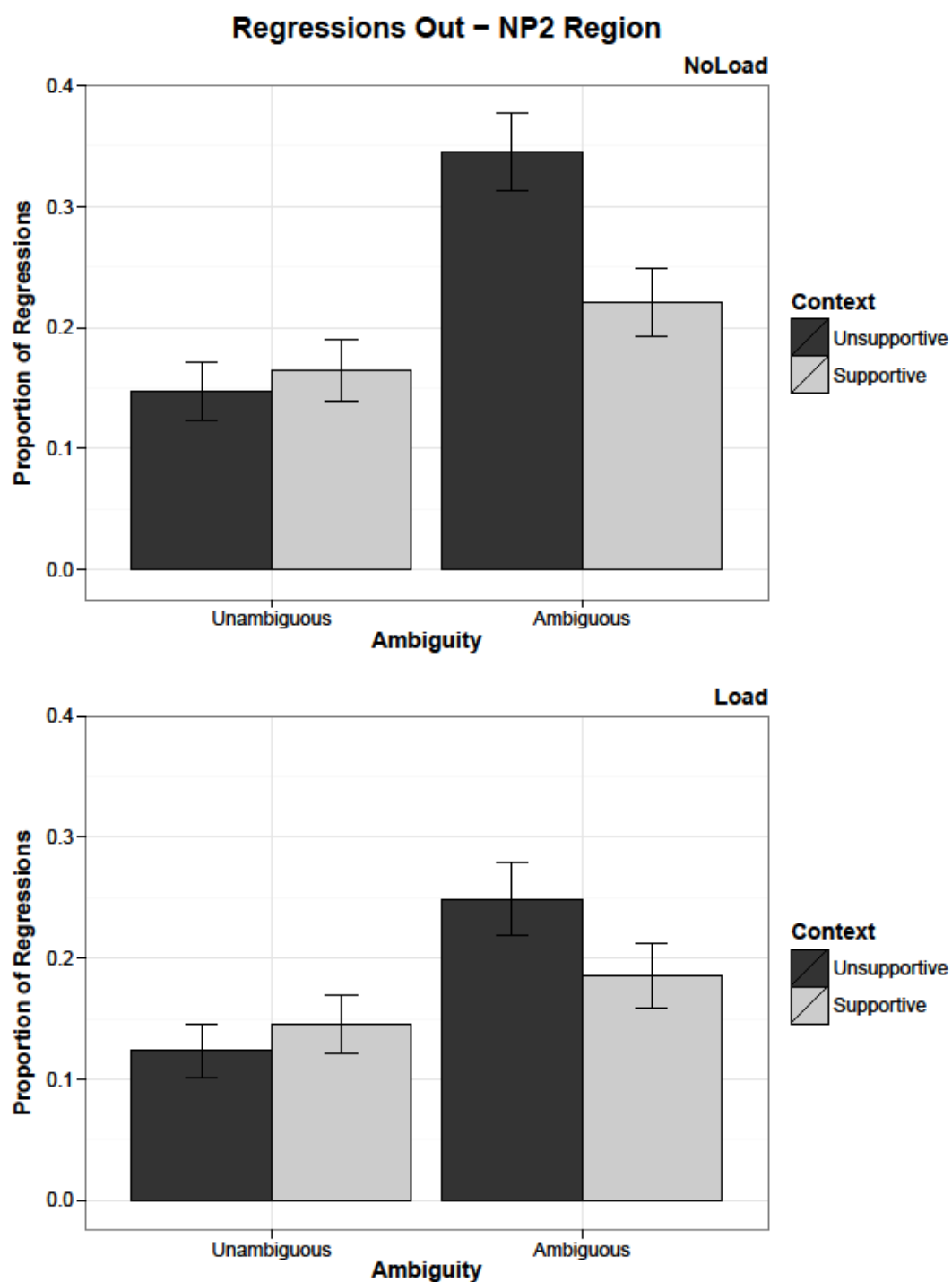


Figure 5.4: Combined Analysis of Experiments 1 & 2. Proportion of Regressions for Regressions Out on the NP2 Region for the No Load and Load experiments. Error Bars display Standard Error. A non-significant Context x Ambiguity x Load Interaction.

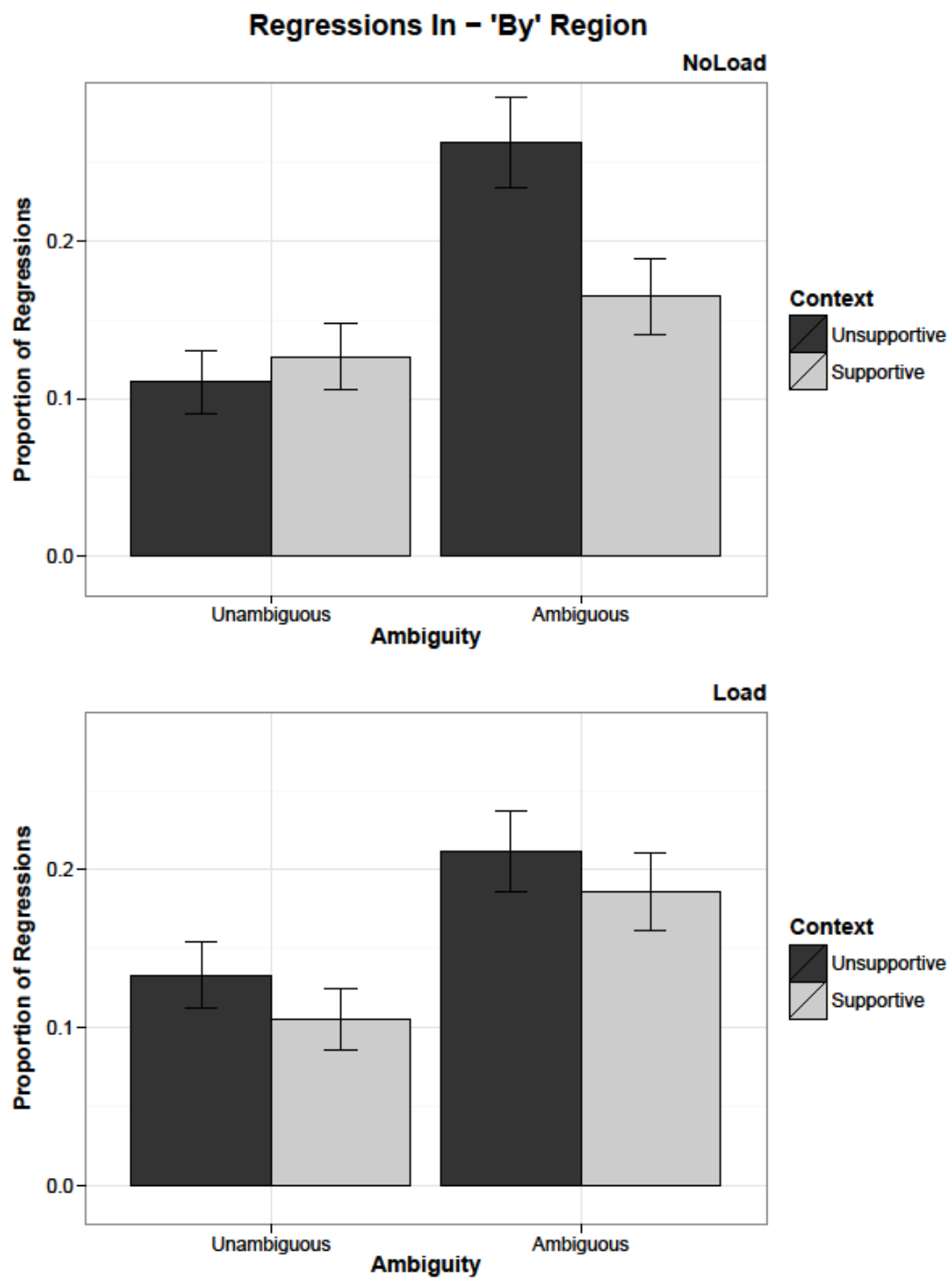


Figure 5.5: Combined Analysis of Experiments 1 & 2. Proportion of Regressions for Regressions In on the 'by' Region for the No Load and Load experiments. Error Bars display Standard Error. A marginal Context x Ambiguity x Load Interaction.

### Other Interactions:

In the sections below we will discuss any relevant interactions involving Context, Ambiguity, and Cognitive Load. In many cases we find interactions between the same factors patterning the same way, just in different measures or regions. Any which pattern differently or are of particular interest will be discussed, but rather than detailing each one of the similar interactions in full, we provide a list of those not discussed in Table 5.3.

To anticipate our results, we find the following interactions:

Context by Load interactions:

#### *First Pass on the Verb*

Participants not under Load show shorter reading times in the supportive context than the unsupportive context, but Loaded participants show similar reading times for both contextual conditions (see Figure 5.6).

Ambiguity by Load interactions:

#### *Probability of Fixation on the Verb*

We see a higher likelihood of fixating the Verb and Verb+By regions for the participants not under cognitive load, but no difference in likelihood for Loaded participants. Assuming that we expect fixations on the disambiguating regions, this implies that Loaded participants are not able to utilise preview and predictive functions to guide their eye-movements in the same way that non-Loaded participants are (see Figure 5.7).

#### *Regressions Out on the Verb*

There is a higher likelihood of regression when encountering an ambiguity for both resource groups. However, when loaded, participants were less likely to regress out when encountering unambiguous sentences (see Figure 5.8).

Table 5.3: Combined Analysis of Experiments 1 & 2 – Interactions which are not discussed in the text.

Interactions	Discussed		Not Discussed	
	Measure	Region	Measure	Region
Context + Load	First Pass	Verb	First Fixation Regression Path	Verb+By Verb
Ambiguity + Load	Probability of Fixation	Verb	Probability of Fixation	Verb+By
	Regressions Out	Verb	Regressions Out	Verb+By

### Context by Load Interaction:

First Pass of the Verb:

In First Pass we observe a Context by Load interaction ( $p = 0.0198 *$ ) on the Verb in which we see longer first pass reading times when participants are not under load and the context is unsupportive (278ms), compared to any other condition (see Figure 5.6) . All other conditions are essentially the same (NoLoad / Supportive = 251ms, Load / Unsupportive = 254ms, Load / Supportive = 250ms). This is an interesting observation as it implies that those participants who are not under Load are able to make use of context to facilitate processing, but Loaded participants gain no assistance from a supportive context or hindrance from an unsupportive one. Instead, it is likely that the greater overall discourse cohesion provided by the context provides the facilitation we see in the NoLoad condition, and that participants which are under Load are reading shallowly and performing integration at the end of the sentence rather than incrementally. Thus when the loaded group appear to be gaining contextual facilitation in both context conditions, they are in fact not making use of context at all, but are gaining what looks like a facilitation effect by reading rapidly and putting off processing until later. This assertion is supported by observations of main effects of Load, which are discussed at the end of this section.



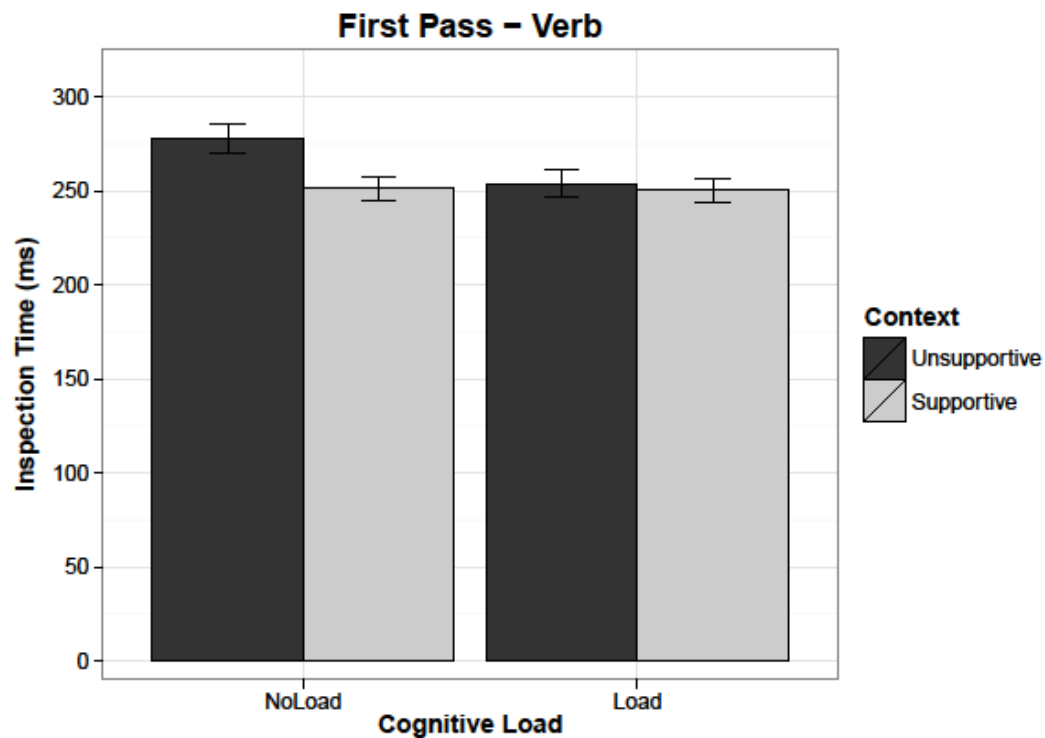


Figure 5.6: Combined Analysis of Experiments 1 & 2. Mean inspection time for First Pass on the Verb. Error Bars display Standard Error. Context x Load Interaction.

#### Ambiguity by Load Interactions:

Probability of Fixating the Verb:

We observe an Ambiguity by Cognitive Load interaction on the probability of fixating the Verb ( $p = 0.026$  \*, see Figure 5.7). When participants are under Cognitive Load we see a higher probability of fixation if the relativizer is not present than if it is (85% vs. 77%). However, while under Load participants show no difference in likelihood of fixation based on ambiguity (Unambiguous = 74% vs. Ambiguous = 74%). The implication of this is that participants are more aware of upcoming ambiguity when they have cognitive resources to spare (i.e. not under Cognitive Load) and make use of that information to guide fixations, whereas those under cognitive load are unable to do so. This raises the question of where the bottleneck lies: is it that participants under load have a reduced forward-span and are unaware of the upcoming ambiguity, or that they their span is unaffected and they are aware of it but are unable to make use of the information? If the latter, is it a case of being

unable to make use of the information at all, or just an inability to do so in time to guide the next fixation? We will address these questions in the Discussion (section 5.4).

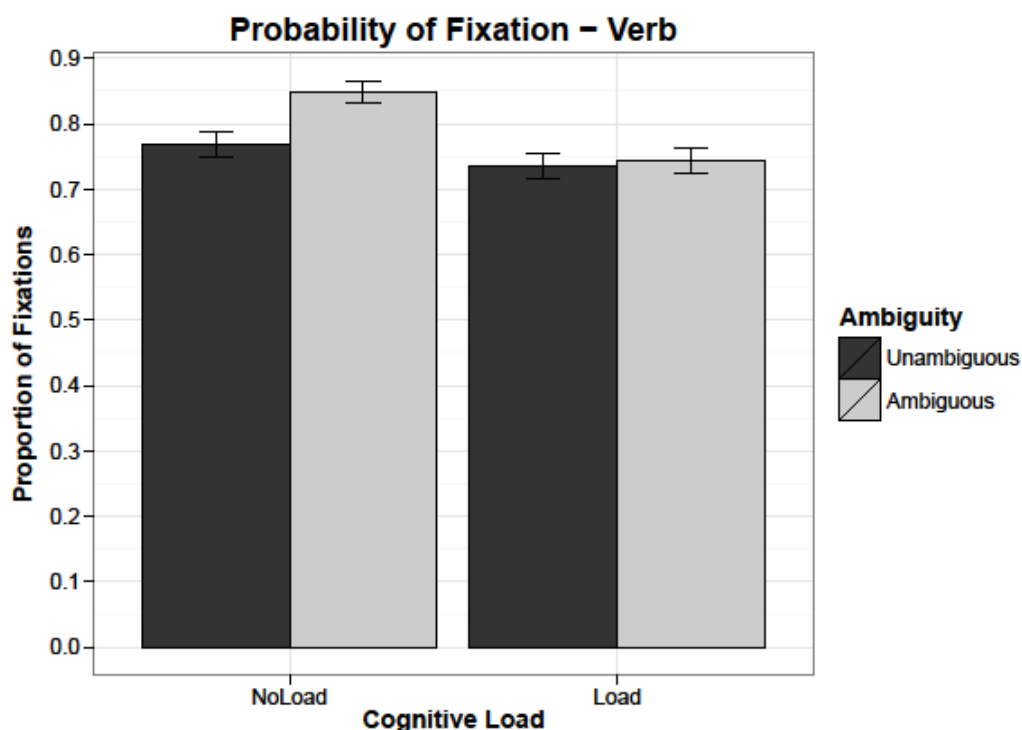


Figure 5.7: Combined Analysis of Experiments 1 & 2. Proportion of Fixations for Probability of Fixation on the Verb. Error Bars display Standard Error. Ambiguity x Load Interaction.

#### Regressions Out of the Verb:

Here we see an Ambiguity by Load interaction in the proportion of regressions out of the Verb ( $p = 0.042 *$ , see Figure 5.8). It shows that when not under Load participants make more regressions out of the Verb region when the relativizer is not present (Ambiguous = 27%) compared to when it is present (Unambiguous = 20%). Similarly, we see the same pattern when participants are under Load, but to a much greater extent: (Ambiguous = 25% vs. Unambiguous = 12%) It therefore seems that while an Ambiguity causes more Regressions Out regardless of Load, a syntactic structure which is Unambiguous reveals an innate bias to not regress out of the Verb or Verb+By regions while under Cognitive Load.

This fits with the hypothesis that Loaded participants are reading more shallowly and quickly, and processing less incrementally than normal.

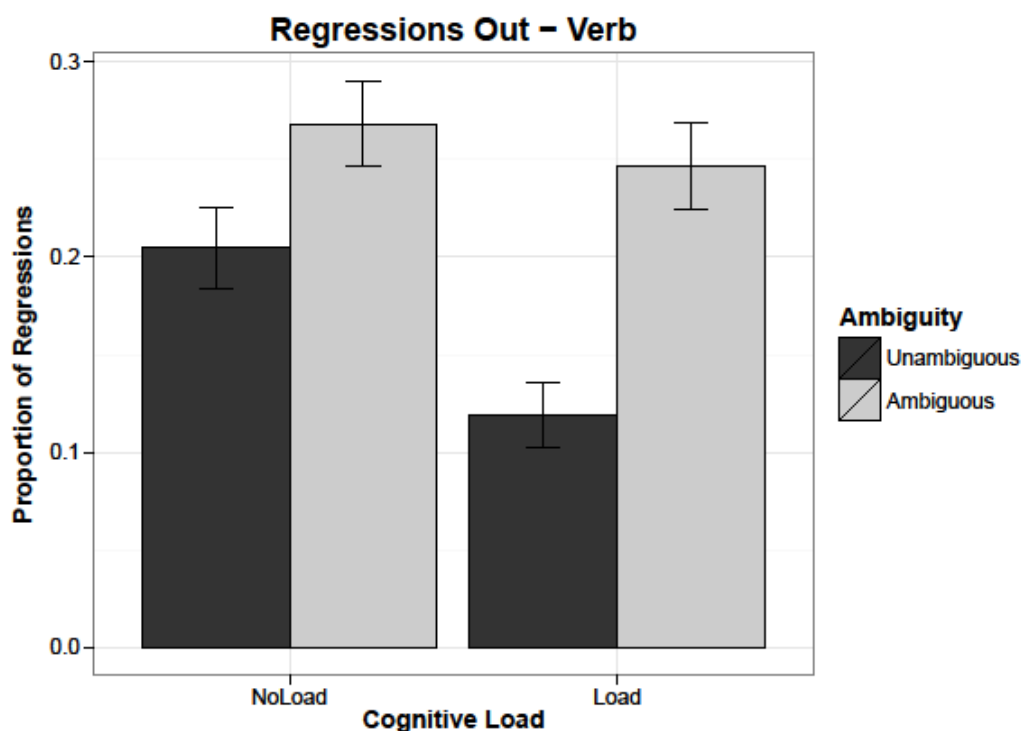


Figure 5.8: Combined Analysis of Experiments 1 & 2. Proportion of Regressions for Regressions Out on the Verb. Error Bars display Standard Error. Ambiguity x Load Interaction.

### Combined Analysis, Summary of Main effects:

The main effects are reported in the tables below. Table 5.4 reports Context effects, while Table 5.5 reports Ambiguity effects, and Table 5.6 reports Load effects.

#### Context:

With the additional power provided by combining the two experiments, we see previously unobserved main effects of Context in the 'by' region (FP, FF), and the NP2 region (RP). In all cases there are longer reading times when the context is unsupportive. Unsurprisingly, the context effect on the First Fixation of the Verb, which patterned differently in each individual experiment, is not observed in either direction when the two are combined. No

other effects were in conflict between the experiments although the marginal probability of fixation effect on the NP1 region observed in the Load experiment, and the marginal Total Time effect on the Verb+By region observed in the NoLoad experiment are both now significant. Conversely, several effects have disappeared as a result of combining the two experiments ('by' region: RP; Verb+By region: FP, FF, SP; Spillover region: TT, Pfix). Overall, as noted in the introduction, there have been severe reductions to the prevalence of Context effects, but not a total elimination. Effects also, in the main, still follow the predicted pattern of facilitation in the presence of a supportive context (see Table 5.4).

#### Ambiguity:

The additional power of combining the two experiments enables us to see two new Ambiguity effects in First Fixation (Verb, Verb+By). Both pattern in the same way as the general trend: Longer fixation durations if the syntax is locally ambiguous). Additionally, the marginal ambiguity effect on the 'by' region for regression path is now observed as significant, whereas the By+NP2 first pass effect which was significant in the NoLoad experiment is now marginal (although patterns the same), and the probability of fixation effect on the spillover region observed in the Load experiment is now no longer significant (or even marginal). Overall, there has been little change to the Ambiguity effects, and the overriding trend is still for an ambiguity to cause reading times and higher proportions of regressions (see Table 5.5).

#### Load:

The Load factor discussed below is simply an indicator of experiment, rather than Load itself, and while it is the most likely variable to cause differences between the groups it is not possible to rule out other variables like overall reading speed. Therefore, we have discussed the effects of the Load factor as if they directly indicate the influence Load, but the reader should bear in mind that other factors may be at play.

As Table 5.6 indicates, early reading measures (FP, FF, RP) appear greatly affected by Load, such that being under load reduces reading times and fixation durations on and around the 'by' region. Fixation measures follow this pattern as well, with the probability of fixating, and the proportion of regressions out from the Verb, Verb+By, and By+NP2 regions, being reduced under Load. There are two exceptions to this, firstly participants under Load spend significantly more time re-reading (SP) the By+NP2 region, and secondly, although not

displayed on Table 5.6, participants under Load make more regressions in to the context region prior to the target sentence, than do those not under Load. Overall it appears that participants who are under Load read more quickly (and presumably more shallowly), skip more often, and make fewer regressions out, but they spend longer re-reading critical areas after having read the whole text, and regress into the context area more often, presumably in an attempt to gain more information or validate their (partial) comprehension.

Table 5.4: Combined Analysis of Experiments 1 & 2 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if Context is **Unsupportive** or **Supportive**. Effects are also marked if **Marginal**.

<b>Context</b>	<b>Levels</b>	<b>NP1</b>		<b>Verb</b>		<b>By</b>		<b>NP2</b>		<b>Spillover</b>		<b>Verb+By</b>		<b>By+NP2</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unsupportive	0.81	334.9	-2.21	265.3	-2.02	241.2	-1.00	279.7	0.47	846.0	-1.71	284.6	-0.70	320.1
	Supportive		379.4	*	250.2	*	213.2		277.7		866.8		26.57		316.8
First Fixation	Unsupportive	-1.83 m	237.8	-0.12	228.3	-2.00	235.2	-0.32	221.1	-0.48	234.8	-0.77	227.2	-0.64	224.0
	Supportive		227.6		221.2		*		219.2		234.0		224.4		220.7
Regression Path	Unsupportive	0.47	419.2	-2.60	423.7	-1.43	386.9	-2.31	411.8	-1.17	1283	-2.38	464.0	-1.71	455.6
	Supportive		483.0	**	375.2		370.7	*	403.3		1245	*	417.1		468.3
Second Pass	Unsupportive	0.60	176.5	0.61	161.4	-1.01	44.4	-0.62	135.1	-1.88 m	278.1	0.03	210.8	-0.91	170.2
	Supportive		208.4		169.3		39.5		128.7		236.8		211.1		159.8
Total Time	Unsupportive	0.77	448.3	-2.24	415.2	-0.66	259.5	-1.07	409.9	-1.61	1116	-2.51	481.6	-1.57	483.6
	Supportive		531.4	*	406.3		255.1		399.7		1105	*	56.4		466.6
Probability of Fixation	Unsupportive	2.12	.70	-3.27	.80	1.06	.16	-0.42	.87	-1.66	.98	-2.59	.87	-0.69	.91
	Supportive	*	.73	**	.75		.18		.86		.97	***	.83		.90
Regressions Out	Unsupportive	-0.86	.06	-0.26	.21	0.00	.19	-1.55	.22	-0.97	.31	-0.64	.21	-0.57	.17
	Supportive		.05		.21		.18		.18		.29		.20		.16
Regressions In	Unsupportive	-3.27	.46	-0.30	.37	-1.80	.18	0.13	.30	0.02	.20	-1.89	.44	0.03	.33
	Supportive	***	.39		.36		.15		.30		.20	m	.39		.33

Table 5.5: Combined Analysis of Experiments 1 & 2 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if the local Syntax is **Ambiguous** or **Unambiguous**. Effects are also marked if **Marginal**.

<b>Ambiguity</b>	<b>Levels</b>	<b>NP1</b>		<b>Verb</b>		<b>By</b>		<b>NP2</b>		<b>Spillover</b>		<b>Verb+By</b>		<b>By+NP2</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unambiguous	-0.19	366.5	4.28	242.7	1.19	221.8	0.36	277.8	-3.73	884.2	5.04	256.7	1.94	312.2
	Ambiguous		347.7	***	272.8		232.6		279.6	***	828.6	***	295.4	<sup>m</sup>	324.7
First Fixation	Unambiguous	-0.26	234.1	2.14	221.2	0.87	218.3	-0.06	219.4	0.57	233.2	2.08	220.1	0.56	220.6
	Ambiguous		231.4	*	234.1		225.0		220.9		235.7	*	231.6		224.0
Regression Path	Unambiguous	0.30	458.2	7.41	361.7	2.42	232.1	5.15	393.2	3.48	1208	8.45	385.4	6.45	429.6
	Ambiguous		444.0	***	437.2	*	434.5	***	423.0	**	1320	***	495.7	***	494.4
Second Pass	Unambiguous	3.82	164.8	8.90	106.6	5.89	28.09	4.32	107.7	3.49	217.7	9.55	137.3	5.60	126.8
	Ambiguous	***	220.1	***	223.8	***	55.88	***	156.0	***	297.2	***	284.6	***	203.2
Total Time	Unambiguous	3.82	474.4	11.64	245.0	3.99	236.7	5.28	382.3	1.40	1101	13.98	385.6	7.92	433.9
	Ambiguous	***	505.2	***	476.5	***	277.9	***	427.3		1119	***	561.5	***	516.3
Probability of Fixation	Unambiguous	-1.73	.73	2.83	.75	1.37	.16	-1.01	.87	0.86	.97	1.67	.84	-2.00	.92
	Ambiguous		.70	**	.80		.18		.86		.98		.86	*	.89
Regressions Out	Unambiguous	0.12	.05	5.04	.16	2.61	.14	5.52	.15	4.54	.26	5.76	.16	5.82	.12
	Ambiguous		.06	***	.26	**	.24	***	.25	***	.35	***	.26	***	.22
Regressions In	Unambiguous	4.31	.38	8.74	.27	5.37	.12	4.30	.26	0.11	.20	9.00	.32	4.66	.29
	Ambiguous	***	.47	***	.45	***	.21	***	.34		.20	***	.51	***	.38

Table 5.6: Combined Analysis of Experiments 1 & 2 - Main Effects of Load. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if participants are under **Load** or **NoLoad**. Effects are also marked if **Marginal**.

<b>Load</b>	<b>Levels</b>	<b>NP1</b>		<b>Verb</b>		<b>By</b>		<b>NP2</b>		<b>Spillover</b>		<b>Verb+By</b>		<b>By+NP2</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	NoLoad	-0.44	355.0	-1.09	263.7	-2.30	238.7	-1.48	290.2	-2.32	922.5	-1.30	285.4	-1.83	335.7
	Load		359.2		541.8		*		267.2		*		266.7		301.2
First Fixation	NoLoad	0.11	230.0	-0.53	229.3	-2.03	229.9	-0.17	220.6	-1.33	239.1	-0.63	227.8	-0.56	224.4
	Load		235.5		226.0		*		219.8		229.8		223.9		220.2
Regression Path	NoLoad	0.06	433.4	-2.38	425.5	-1.94	401.3	-2.11	422.6	-1.87	1295	-2.61	475.7	-2.21	482.4
	Load		468.9		*		<sup>m</sup>		*		<sup>m</sup>		*		*
Second Pass	NoLoad	1.41	173.6	0.89	155.3	0.12	41.59	2.39	108.6	1.64	224.0	0.59	203.0	2.23	138.4
	Load		211.3		175.1		42.38		*		290.9		218.9		*
Total Time	NoLoad	0.37	463.9	-0.28	401.9	-1.00	259.3	-0.15	392.0	-1.73	1140	-0.89	472.4	-0.04	464.1
	Load		515.7		419.5		255.3		417.6		1080		474.6		486.1
Probability of Fixation	NoLoad	-0.83	.73	-2.25	.81	-0.32	.18	-2.16	.90	-0.64	.98	-1.90	.88	-2.65	.94
	Load		.70		*		.17		*		.97		<sup>m</sup> .83		**
Regressions Out	NoLoad	-0.01	.06	-2.00	.24	-1.03	.21	-1.41	.22	0.36	.29	-2.10	.24	-0.84	.18
	Load		.05		*		.16		.18		.31		*		.16
Regressions In	NoLoad	0.32	.42	0.91	.34	-0.11	.71	1.64	.27	1.53	.17	0.74	.40	1.58	.30
	Load		.44		.38		.16		.33		.24		.43		.36



## 5.4 Discussion

In the introduction to this chapter we detailed the main effects and interactions which appeared to differ between the NoLoad and Load experiments. Particularly, we highlighted two Context by Ambiguity interactions, one in Regressions Out of the NP2 region, and one in Regressions In to the disambiguating 'by' region. In both cases the interaction was present in the NoLoad experiment, but not observed in the Load experiment. It appeared that under normal circumstances participants were able to make use of context to help resolve syntactic ambiguities, but while loaded their ability to do so was much reduced.

We examined these interactions statistically and found that there was no reliable difference between the two groups for the Regressions Out interaction. For the yoked Regressions In interaction we observed a marginal difference between NoLoad and Loaded participants in the direction that we anticipated, but it is unclear whether this three way interaction would become significant with the addition of more experimental power or not.

Conducting a statistical meta-analysis of the two experiments has also suggested that being under Load affects reading behaviour by generally speeding fixations, but not by increasing the proportion of regressions out of any particular region. This is assumed to be a speed / accuracy trade off in which processing depth is sacrificed for the sake of finishing the reading task and recalling the digits sooner. It has also been suggested that Load interacts with Context and Ambiguity separately:

We have reported several different manifestations of a Context by Load interaction in the early reading measures on and around the Verb region. Each indicates that, regardless of the presence or absence of an ambiguity, NoLoad participants make use of a supportive context to speed processing, while Loaded participants do not do so. The question of whether they are not doing so because they are unable to, or because they are rushing due to task demands remains in debate.

Similarly, we observe Ambiguity by Load interactions on and around the Verb region for both Probability of Fixation and Regressions Out. For the Probability of Fixation interaction, we observe Loaded participants having no greater likelihood to fixate the Verb or Verb+By regions depending on Ambiguity, while participants not under Load had a higher probability of fixation if the upcoming verb / Verb+By region was Ambiguous. We considered whether

the necessary difference between the Unambiguous and Ambiguous sentences (the presence / absence of a 7 character 'who was' relativizer region), in combination with the general tendency for Loaded participants to skip more (see the Number of Fixations measure in the introduction to this chapter) might have generated such a pattern of data. We reject this idea as it seems unlikely that such an interaction of factors would produce this pattern. If it was the case that both Loaded and non-Loaded participants generally fixated the relativizer, but that the higher tendency for Loaded participants to skip regions caused them to be more likely to skip the Verb / Verb+By regions, then we would expect to see variation (dependent on ambiguity) in the Loaded participants' probability of fixation. We do not. Therefore the question raised in the results section still stands: what makes participants under Load not react to the ambiguity. We suggest four possibilities: If Loaded participants do not use preview to guide their eye-movements is it because they have a reduced forward span (from the normal 15 characters (McConkie & Rayner, 1975, 1976)) and do not notice the upcoming ambiguity? Do they not have the resources to adapt their fixation behaviour? Or are they unable to adapt their fixation plan in the time allowed (programming and launching a saccade takes time, and it is possible that the reduction in cognitive resources increases the amount of time required for that process to complete)? Finally, do they simply not care enough about the correct interpretation of the sentence to worry about the making sure they fixate on the Verb / Verb+By regions? Our experiment does not distinguish between these accounts but it seems unlikely that cognitive loading impinges on the basic properties of the perceptual span. It is more likely that a reduction in resources prevents the adaption of eye-fixation behaviour, or that participants are not concerned by up-taking the information in the Verb / Verb+By regions. The latter explanation is supported by the general trend toward skipping regions, and speeding / shallowing of processing, although comprehension question answers met an 80% threshold for both experiments so comprehension did not observably suffer.

For the Regressions Out interaction (Ambiguity x Load) in the same regions we see further evidence that Loaded participants have different reading behaviour to those not under load. Specifically, that they make generally fewer regressions out of the Verb / Verb+By regions, but importantly, the proportions by which they do so in the absence of an ambiguity are much reduced, revealing an innate bias to not regress out which supports the speeding / shallowing hypothesis.

Overall the results of this combined analysis show that reading behaviour is influenced in various ways by maintaining a cognitive load. However, contrary to the predictions of the capacity constrained resource model and our assertions at the start of this meta-analysis, several observations indicate that the people might be able to use context to help with the processing of / recovery from a syntactic ambiguity. Firstly, the lack of a three way interaction between Context, Ambiguity, and Load in the Regressions Out measure for the NP2 region: This interaction patterns the same way in the Load condition as for the NoLoad condition, but with a general reduction in the total number of observations of regressions. Secondly, the Regressions In interaction in the Verb region observed in the NoLoad experiment, but not observed in the Load experiment hints towards a three way interaction between Context, Ambiguity, and Load, but the pattern makes little sense (reversing the context effect on the unambiguous condition when under Load) and the overall interaction is only marginal. Thirdly, there is an observable two way interaction between Context and Ambiguity in the NP2 region for Regressions out, which mirrors that seen in the NoLoad experiment. Previously, that was taken as evidence of contextual facilitation of recovery from a garden path, so for it to be replicated in the combined analysis, without interacting with Load, indicates that Load did not eliminate or significantly reduce the use of context when processing syntactic ambiguities.

## Chapter 6: End of Syntax Section - Summary and Discussion

We have conducted two experiments and a meta-analysis investigating people's ability to utilise context to facilitate the processing of syntactic ambiguities while under cognitive load. The first experiment established a baseline of how people performed while not under load, the second was a replication of the first but with the addition of a titrated memory load in the form of a string of digits to be maintained in memory during the reading phase, and the meta-analysis provided a statistical comparison of the two.

These experiments were designed to inform the question of whether there exists a single resource pool (Just & Carpenter, 1992) or a fractionated resource pool with components specific to certain language processing functions (Caplan & Waters, 1999). The SR (Single Resource) theory posits that there is an overriding resource pool from which all cognitive functions draw, and that this system is, architecturally speaking, an interconnected, interactive unit. However, as the resources are depleted the system begins to become more modular, with there being too few available resources for different language processes to interact with one another. In this way the SR becomes functionally fractionated. On the other hand, the SLIR (Separate Language Interpretation Resource) theory holds that there are at least two architecturally separate pools of resources dedicated to language processing. The Interpretive resource allows for automatic language processes, such as syntactic processing, lexical access and the resolution of causal inferences. The Post-Interpretive resource allows for more conscious language processing, such as the resolution of garden paths, and the memory of lists.

The SR theory therefore predicts that under cognitive load there should be an increase in modularity (i.e. a reduction in the ability to bring in high level contextual information to inform syntactic processing and reanalysis), while the SLIR theory predicts no influence of loading on normal syntactic processing, but a reduction in people's ability to bring in contextual information to assist with recovery from a garden path.

In the first experiment we observed that people regressed out of the NP2 region and into the 'by' region (the baby frightened *by* the flames) significantly less when they encountered an ambiguity if the context had not supported the relative clause interpretation of the syntax, compared to when it had, whereas the presence of a supportive Context made no

difference when the sentence was not ambiguous. The presence of a main effect of ambiguity in the preceding region (first pass) suggested that participants were encountering the ambiguity, getting garden-pathed, and then utilising the context to facilitate the recovery from the garden-path. As mentioned, it is not the intention of this thesis to provide evidence pertinent to the syntax-first / parallel competition / parallel race models of parsing, so we will not discuss how our findings impact that debate in any detail. However, we direct the reader to the work of Clifton & Staub (2008) and Farmer et al. (2007) for good reviews of this discussion. For the purposes of this discussion we will couch our findings in a syntax-first perspective, as that most simply explains the data (note that both the parallel competition and parallel race models can also explain the data, but not as simply).

The second experiment was identical to the first except in that participants were under a titrated memory load designed to reduce cognitive resources. We observed both the previously mentioned interactions disappear. The Regressions Out interaction patterned the same but had fewer overall observations of regressions and was not statistically significant, while the Regressions In interaction was also non-significant but patterned differently in a non-theoretically interesting way. This, combined with a large apparent reduction in the influence of Context around the critical regions led us to conclude that people were less able to make use of context to assist in parsing sentences, and particularly ambiguities, when under cognitive load. Interestingly, the participants in this study did not perform in quite the same way as the Low Span readers of MacDonald et al. (1992) who, according to the results of comprehension questions, did not always make an attempt to process the ambiguity correctly, or revise their incorrect interpretation upon disambiguation. In our study all participants met an 80% threshold for correct comprehension question answers so we know that they were trying, with a high rate of success, to achieve the correct interpretation. If we assume that our resource manipulation has been successful then our participants should be comparable to those with a low working memory span. Therefore, the overall capacity of the Low Span and resource limited readers is similar, but those who are being experimentally manipulated maintain the experience and ability to recognise and correctly process (possibly through revision) syntactic ambiguities, while those of a naturally low span do not. If this is so then this observation lends support to the suggestion that it is not the total amount of capacity available which dictates the processing ability of the system, but rather the skill and

efficiency of the processing mechanisms, as suggested by MacDonald & Christiansen (2002) in their 'experiential account' (see Literature review, section 1.3).

In contrast to the findings of the second experiment, the statistical meta-analysis between the two experiments showed few reliable differences between the two groups, although there are noteworthy interactions of Load and context in the Verb and Verb+By regions for the early reading measures (First Pass, First Fixation, and Regression Path) which appear to show participants utilising context during parsing when not loaded, but not when loaded (regardless of ambiguity). Importantly, the Regressions Out Context by Ambiguity interaction did not interact with Load. This suggests that, contrary to our initial predictions, people under cognitive load are not able to utilise context to facilitate the parsing of sentences, but are able to so do during the processing of syntactic ambiguities.

Our findings are at odds with both the SR and SLIR accounts. As mentioned above, the SR account would predict that, as cognitive resources are reduced, the interactivity of the system becomes lessened and processes like syntactic parsing, ambiguity resolution, and contextual processing should no longer interact with one another. We should have observed a difference between the two Load conditions such that contextual information was no longer applied in the resolution of the ambiguity, or in the general parsing of the sentence. The SLIR account would predict two things: Firstly, a similar stance on modularity with regard to syntactic reanalysis (no contextual information being used to assist in the recovery from a garden path), but secondly, contextual information being utilised in non-ambiguous sentence processing regardless of Load.

We see no support for the shared prediction that context should not be used by Loaded participants in circumstances of syntactic ambiguity (Context by Ambiguity interaction in Regressions Out of the NP2 region, which did not interact with Load, indicating no difference in the use of context to facilitate recovery from a garden path between non-Loaded and Loaded participants). And mixed evidence pertaining to the use of context in normal parsing: In support of the SR account we see FP, FF, and RP Load by Context interactions in the Verb and Verb+By regions suggesting that participants that were not under Load did make use of context in normal parsing, but that those that were did not. But in support of the SLIR we see main effects of Context such that longer reading times were observed in the critical regions when the context was not supportive, implying that contextual facilitation was not influenced by load.

Therefore, neither theory fully accounts for our data. There are three possibilities as to why our data does not fit well with either account: Our initial observation of the differences between experiments was correct but the experiments do not have the power for the effect to be statistically reliable, the task demands of maintaining / recalling the digits caused participants to develop abnormal strategies which influenced reading behaviour and obfuscated the results, or our digit span manipulation did not reduce resources appropriately. We will discuss these in turn.

Making a non-statistical comparison between experiment 1 and experiment 2 (NoLoad vs. Load) we can see that the majority of the main effects of Context present in normal processing, and the important Context by Ambiguity interactions, disappear under Load. It is possible that our experiments simply did not have the required power for the three way interaction to be reported as statistically significant. If running more participants, or otherwise increasing the power of the experiment, had made the differences significant then our findings would provide support for the SR account: A reduction in the overall interactivity of the system such that context could not be used to either to speed normal language processing, or facilitate ambiguity resolution. It is worth noting that the level and type of load here does not eradicate the use of context completely, only reduce it. Even if more power had made the differences between experiments significant, there were still some instances in the Load experiment in which a supportive context speeded processing, and, notably, the Regressions Out interaction, while not significant in the Load experiment patterned the same in both experiments (with a major reduction in the overall number of regressions for those under Load). This implies that, while loaded, people are much less able to utilise the context in normal reading and in ambiguity resolution, but that they are still able to do so to a reduced degree.

We have noted several times that the results seem to indicate that participants are reading more quickly and more shallowly when under cognitive load. This could be a genuine result of the loading manipulation (in that a reduction in cognitive resources limits the depth of processing during the bulk of reading, and offsets it towards the end of the sentence in an exaggerated wrap-up effect). However, it is also possible that this is simply a strategy devised by participants to rush to the end of the sentence so that they can offload the digits as quickly as possible. The result of either option would be a reduction in fixation durations in the body of the sentence, increased skipping throughout, and longer fixation durations

on later regions and second passes, all of which we observe. Whether participants are using a strategy or not is hard to discern from the data we have, but we were careful to take steps to prevent participants from engaging in consciously non-normal reading behaviour. We stressed the importance of focusing both on the reading and digit recall components of the task, and participants were made aware that they were required to get both components correct. Any data that failed to meet the threshold on either task was discarded to ensure that only data from those participants who were dividing their attention appropriately was analysed. With this in mind, we believe that this pattern of results reflects a direct effect of our Load manipulation, rather than a strategy. The question then raised is: Did our digit load manipulation truly deplete the resources we intended it to?

There has been much debate in the literature as to whether or not digit span (the number of digits in a digit string that a person can encode and recall accurately and in the correct order) is an appropriate measure of working memory. The general conclusion is that it only reflects the storage component of the working memory system, and does not capture the operational component in the same way that the Reading Span does (see Literature Review, section 1.4.2 for discussion). However, working memory is defined by Just & Carpenter (1992) as functionally flexible operational space in which storage and operational processes share resources. It seems logical then that by engaging the available resources in storage (with an extrinsic memory load), operations such as language processing which require both operational and storage components, should see their overall resource availability reduce as resources that would usually be assigned to performing operations, are tied up in providing additional storage. While the statistical analysis does not support the SR accounts' prediction of capacity constrained modularity, or the SLIR accounts' specific prediction of disruption in context use during syntactic reanalysis only, we do observe interactions of Load with Context and Ambiguity individually that lead us to believe that our manipulation was having the desired effect. For example: The above mentioned early reading measure Load by Context interactions (FP, FF, RP on the Verb and Verb+By regions) show participants making use of context when parsing sentences (regardless of ambiguity) when not under load, but not when Loaded. We believe that the interpretations of the statistical and non-statistical comparisons discussed above, both indicate that our Load manipulation is successfully limiting the availability of the storage and operational resources required to perform parsing and syntactic ambiguity resolution.



Out of the three possibilities which we presented as potential reasons for our results not fitting with either theory, we feel that the most likely candidate is that our experiments lacked sufficient power to show statistically the differences in context use between non-loaded and loaded participants. We believe that with more power our findings would have indicated support for the Single Resource account, as all facets of contextual facilitation appeared to be affected, and some of these were significant, while others were not.

Potentially, one further issue reducing the clarity of these results is whether or not syntax is being processed first without contextual input and repair is occurring later on (serial syntax-first models), or if multiple syntactic representations are being maintained and the ambiguity cost we observe in down-stream regions is associated with the selection of a dispreferred interpretation (parallel models). We remained agnostic to this issue as it is beyond the scope of the thesis to weigh in on that argument, and our experiments were not designed to provide support one way or the other. However, while it should not particularly impact our results in terms of distinguishing between the SR and SLIR accounts, it is possible that if syntactic processing has some privileged status above and beyond other language processes our manipulation might not be as impacting as it could be. In the next section we utilise the same methodologies to examine another language process which does not have such potential for privileged status: Lexical Ambiguity.

### Section 3: Cognitive Load, Context, and Lexical Ambiguity Resolution

Homonyms are words that have multiple meanings but which are written and pronounced the same. This leads to difficulties in language processing as when the reader / listener encounters these words they cannot be sure which meaning is the correct one. The example below exemplifies this:

*The boy visited the bank last Tuesday.*

In this example the word *bank* is ambiguous such that it could mean a financial institution or a river bank. Without any further information the reader cannot be certain which interpretation to select. However, in the case of *bank* the two meanings of the word are not equal in terms of frequency of usage; the financial sense being much more frequently used than the river sense, among the majority of the populations that psycholinguists study. This use of frequency leads people to be more likely to interpret *bank* as the financial institution when it is presented in isolation (MacDonald et al., 1994; Rayner & Duffy, 1986; Swinney, 1979). We refer to these ambiguities as *biased* ambiguities.

However, some words are ambiguous yet do not have a particularly dominant sense in terms of frequency. We refer to these ambiguities as *balanced*. It is worth noting that the frequency of a word's senses varies depending on the population which is being examined. The sentence below is taken from Rayner & Duffy (1986) and uses the word *pitcher*: To the local population of Rayner and Duffy's research centre (Amherst, Massachusetts) *pitcher* can either refer to a large jug, or a baseball player.

*Of course the pitcher was often forgotten.*

To the population Rayner and Duffy were testing, *pitcher* is a balanced ambiguity, so frequency information is less useful and people are less able to make use of it to assist with the resolution of such an isolated sentence. In our local population (Edinburgh, Scotland) baseball is not such a pervasive sport and *pitcher* would be considered a biased ambiguity. This issue of localisation is discussed further in Section 7.1

What mechanisms then, are considered to underlie the processing of these different types of ambiguity? In 1986, Rayner and Duffy examined lexical access of biased and balanced ambiguous words with a disambiguating context which was only available downstream of the ambiguity. They compared the eye-movements of people reading balanced and biased ambiguities to unambiguous frequency matched controls, as can be seen in the examples given below. Note that the given downstream context supports the less frequent interpretation of the ambiguity, therefore lexical misinterpretations and reanalysis are likely to occur once the disambiguating information is reached.

**Biased ambiguity:** I know the *pen* is too tiny for an elephant.

**Unambiguous control:** I know the *zoo* is too tiny for an elephant.

**Balanced ambiguity:** He found the *coach* was too hot to sleep in.

**Ambiguous control:** He found the *cabin* was too hot to sleep in.

They found an interesting pattern of results such that: For biased ambiguities no disruption was observed on the ambiguity itself compared to the unambiguous control, however, there was significant disruption when the disambiguating information was encountered.

For the balanced ambiguities there was slow down on the ambiguous word compared to the unambiguous control, but much less disruption on the region containing the disambiguating information than was seen for the biased ambiguities.

Duffy, Morris, & Rayner (1988) discuss these findings in terms of the models of lexical ambiguity resolution. They assert that the findings fit with the Exhaustive Access model, in which all senses of an ambiguous word are accessed and, subsequently, a single option is selected. Critically, the order and speed of access is related to the frequency of each sense. Biased ambiguities (which have one sense of a much higher frequency than the other) see their more dominant sense accessed first, and selected before the less dominant sense can become fully accessed. This means, similar to a syntactic garden path, a selection is made based on inherent frequency bias (so no delay is observed compared to frequency matched controls), which is then proved either right or wrong by disambiguating information downstream. If it is proven right then there is no cost of reanalysis as no reanalysis occurs, but if it is proven to be incorrect then reanalysis takes place and disruption is observed on the disambiguating information. Balanced ambiguities on the other hand, see their senses

accessed simultaneously which is costly comparative to unambiguous controls, not only because of the cost associated with fully accessing two words compared to one, but also due to the testing of each of those accessed words against the downstream context in order to make the correct selection.

But what happens if the context precedes the ambiguity? Research using a cross-modal priming paradigm (Onifer & Swinney, 1981; Patrizia Tabossi, Colombo, & Job, 1987), eye tracking (Rayner & Frazier, 1989), and ERP (Swaab, Brown, & Hagoort, 2003) has shown that even if a preceding context has strongly implicated one interpretation over another, both interpretations are immediately accessed. However, during the subsequent selection phase, the supported interpretation will receive more activation than the non-supported one and be selected for further processing. The non-selected interpretation will decay or be actively inhibited. See Seidenberg, Tanenhaus, Leiman, & Bienkowski (1982) for an overview of the early research in this area.

Duffy et al. (1988) expanded on Rayner & Duffy (1986) to examine what happened when the disambiguating context *preceded* the biased and balanced ambiguous words. They created sentences which contained either a biased or balanced ambiguity, or a frequency matched control, and had a disambiguating context which could be presented either before or after the ambiguity / control. In the examples on the next page, the context is italicised, the ambiguous word is underlined, and the frequency matched controls are presented in parenthesis.

Duffy et al. discuss three models which describe the mechanisms by which a preceding context may influence lexical ambiguity resolution. The first is the Selective Access model, in which a preceding context provides enough information for the system to only access and select the appropriate sense of the word. The other meanings of the word are not accessed at all. The second and third are two types of Exhaustive Access model. The first is the Autonomous Access model in which a preceding context has no influence on the access or selection of senses of an ambiguity unless it contains a word which specifically primes one sense of the word but not the other. The second is what they term the Reordered Access model. The reordered access model assumes that a preceding context influences the access of the sense it supports in such a way that it makes it more available for selection than the other sense(s).

- a. *When she finally served it to her guests, the port (soup) was a great success.*

*Biased ambiguity, Context before*

- b. *Last night the port (soup) was a great success *when she finally served it to her guests.**

*Biased ambiguity, Context after*

- c. *Because it was kept on the back of a high shelf, the pitcher (whiskey) was often forgotten.*

*Balanced ambiguity, Context before*

- d. *Of course the pitcher (whiskey) was often forgotten *because it was kept on the back of high shelf.**

*Balanced ambiguity, Context after*

Duffy et al. replicated the Rayner & Duffy (1986) findings (above) and extended them such that: A preceding context increased the disruption on the ambiguous word for biased items, and eliminated the disruption on the ambiguous word for balanced items. The increase in processing difficulty of the biased ambiguity is explained by the context being designed to support the less dominant meaning of the word. Thus, by promoting the subordinate meaning, the context creates a situation in which the two interpretations become balanced, and it takes longer for selection to occur. The reverse happens when a context promotes one sense of a balanced ambiguity: The promoted sense achieves selection much more quickly than its competitor. In this way a biased ambiguity behaves much more like a balanced ambiguity, and vice versa. Duffy et al. claim that these findings support their Reordered Access model and offer two options as to how one might conceptualise the

variables in the model working. One way is that sense frequency and context interact and influence the amount of time it takes for a specific sense to be available for post-access selection, and the first sense to reach threshold is selected regardless of the state of the other. The other is to consider that frequency and context provide units of evidence to support one interpretation or another; when one sense is clearly more supported it is selected relatively quickly, but if both senses are equally well supported it takes more time for selection to occur. Duffy et al. point out that “two simultaneously available meanings may divide capacity since some consideration must be given to both meanings and their relationship to the earlier sentence context” (Duffy et al., 1988, p.442) and this resource and capacity perspective has been investigated by Miyake, Just, & Carpenter, 1994.

Miyake et al. (1994) frame the discussion of lexical ambiguity in terms of a working memory capacity restricted model and approach the issue of maintaining multiple interpretations of an ambiguous word from an individual differences perspective. Here we will briefly summarise that which we have so far discussed about lexical ambiguity, and place it within in the framework of Miyake et al.’s model.

When the Ambiguity precedes disambiguating context:

Both interpretations of the ambiguous word are activated on encountering the ambiguity, with their relative frequencies guiding the level of activation that each interpretation achieves (the higher the frequency the more initial activation assigned). The interpretations are elaborated in parallel as evidence is accrued for each one. The greater the weight of evidence in support of an interpretation, the more activation it receives. These parallel elaborations require working memory resources to maintain, and once one interpretation reaches a threshold of activation it is selected and the other is inhibited. If neither achieves threshold soon enough, limitations on the working memory capacity of the reader will cause the weaker interpretation to decay and eventually be discarded once it falls below a minimum activation threshold. High span readers are able to maintain multiple parallel representations at a greater level of elaboration for longer than Low span readers. Biased ambiguities, therefore, start with a much higher level of activation assigned to the high frequency dominant sense than the low frequency subordinate sense, and so the dominant sense is likely to achieve threshold much more quickly. High and Low span readers should find this equally easy to process as the maintenance of multiple parallel interpretations is not such a critical factor unless it turns out, upon disambiguation, that the dominant sense

was incorrect. Upon disambiguation in favour of the subordinate sense, High and Low span readers should show differences, however, as High span readers have the capacity to maintain the subordinate interpretation and so have a head start on recovering the error, whereas Low Span readers will have disregarded the subordinate sense interpretation entirely and will have to undergo full reanalysis. Balanced ambiguities, on the other hand, start with both senses approximately equal in terms of frequency and thus activation. Unless there is strong disambiguating information presented shortly after the ambiguity, readers will be required to maintain and elaborate multiple parallel representations for much longer. High span readers should be better equipped to manage this task than Low span readers and so differences between the two groups should manifest regardless of the outcome of disambiguation.

#### When the Context precedes the Ambiguity:

A context will provide support for one of the potential interpretations in the form of activation. When added to the amount of initial frequency based activation applied to an interpretation, this may be enough for an interpretation to reach the selection threshold and for one sense to be selected and the other inhibited. In the case of biased ambiguities, a context in support of the dominant sense is likely to do this, whereas a context in support of the subordinate sense is likely to balance out the two senses and require the reader to maintain and elaborate two interpretations while further evidence is accrued. Unless the context is so strong that it does not just balance out the initial frequency bias of the dominant sense but instead overcomes it and provides immediate selection, this situation then parallels the 'Balanced ambiguity in the case of no prior context' situation described above. In the case of balanced ambiguities the context is likely to be strong enough to bring either sense to a high enough level of activation to provide immediate, or very quick selection. Therefore, with a balanced ambiguity and a preceding context, High and Low span readers should show little difference.

Using a self-paced reading moving window paradigm and three span groups (High, Medium, and Low) Miyake et al. tested their model's predictions of how readers of differing working memory capacity deal with biased and balanced ambiguities the disambiguating context appears after the ambiguity. Their findings are too multitudinous to describe in detail here but the synopsis is that they fit well with the predictions made by the model outlined above: For biased ambiguities, lower span readers exhibited more difficulty when

encountering the disambiguating information than did higher span readers. Conversely, for balanced ambiguities, even lower span readers were able to maintain both interpretations successfully and as such showed no disruption regardless of whether the disambiguating context eventually supported the dominant or subordinate sense.

These findings, Miyake et al. argue, support the idea that readers always activate all meanings of an ambiguity, and that the level of activation is modulated on the relative frequencies, and contextual support for each meaning. Importantly, readers with limited working memory resources struggle to maintain the activation required to support these multiple interpretations, and so disregard the less activated interpretation more readily than readers with higher working memory capacities. This assertion has found support more recently from fMRI studies which show both a greater intensity of activation in particular brain regions, and activation in brain regions not usually associated with language processing, during lexical ambiguity processing, which are modulated by the working memory span of the participant. Space constraints prohibit us from engaging with the neurological literature but we direct the reader to Mason & Just (2007) for a capacity constrained perspective; Ihara, Hayakawa, Wei, Munetsuna, & Fujimaki (2007) for an examination of selection / integration; and Grindrod & Baum (2003, 2005) for a hemispheric localisation of frequency / contextual influences.

In the upcoming chapters of this section we will extend the capacity oriented work of Miyake et al. (1994) by investigating the role of cognitive resources in the processing of lexical ambiguities under a preceding context. Our principal question (as well as methodology) remains the same as in the Syntax section. We will be using a titrated extrinsic memory load (digit span) to experimentally limit the resources available to participants as they read sentence which are prefaced by either a neutral or supportive context.

We started from the re-ordered access model and the simple question of whether people under an extrinsic memory load were able to make use of contextual information to assist with the disambiguation of lexical ambiguity. We considered using biased ambiguities to inform this question but in the framework of the reordered access model the predictions these make are unclear:



With biased ambiguities, assuming context was available, two possibilities would occur: The first is that contextual support for the subordinate meaning would balance the ambiguity by increasing the likelihood of selection of the subordinate sense to equal the level of the dominant sense. The second is that contextual support for the subordinate meaning would be strong enough to *overbalance* the ambiguity by increasing the likelihood of selection of the subordinate sense to much greater than that of the dominant sense, and causing immediate selection of the subordinate meaning. Unfortunately, it is difficult to predict if a context which supports the subordinate sense would balance or overbalance an ambiguity and the two scenarios have differing predictions. In the case of immediate selection we would expect to see no difference in reading times compared to an unambiguous control but, in the case of the balancing of the ambiguity, we would expect to see increases in reading times as processing load increased (with the maintenance of multiple interpretations).

By comparison, if the context was not present we would expect to see longer reading times due to the maintenance of multiple interpretations (which may decline over time as evidence for the subordinate meaning fails to manifest). This would be difficult to distinguish from the scenario of contextual use in which context balanced the ambiguity, but, as has been mentioned above, it is difficult to predict whether a context would create the balanced situation, or if it might overbalance the ambiguity. A third scenario is possible: immediate selection would also occur if the context supported the dominant meaning, but this would also be hard to distinguish from when no context was present or being used, as readers may just take their frequency bias as sufficient evidence and select the more dominant meaning very quickly. The uncertainty of the predictions of biased ambiguities led us to consider using balanced ambiguities, the predictions from which are much clearer:

If a context is present and being utilised then there is only one anticipated outcome: facilitation. Reading times will be decreased compared to non-contextually mediated controls because the increase in support for one of the interpretations (regardless of dominance) will create speeded selection. The slower that selection occurs the longer the reader must maintain two interpretations and the associated additional load will increase reading times.

Below we put balanced ambiguities in the context of Miyake et al.'s capacity constrained model as a rough guide to how we anticipate participants undergoing an extrinsic memory Load to behave:

High span readers (those not under cognitive load) should be able to maintain two interpretations with little difficulty so should not struggle regardless of the presence of a preceding context. Low span readers (those under cognitive load), on the other hand, should display the stress of maintaining two interpretations in the regions of text following the ambiguity if a context is not present, whereas if a context is present they should show similar processing to the high span readers as only one interpretation needs maintaining.

In the following chapters we will report an online experiment (to generate localised materials), two eye-tracking experiments (one where the participants were not under load, and in which they were), a statistical meta-analysis of these two experiments, and a further eye-tracking experiment utilising a within subjects design for the load factor.

## Chapter 7: Generation of Lexical Items and Materials (Including an Online Experiment)

### 7.1 Introduction to Materials

As has been mentioned above, the relative frequency of words differs by the population tested. For example, in parts of America *Pitcher* is a commonly used word to refer to a jug or a position in baseball, but in the UK this is not the case and both the overall frequency, and the relative frequency of each meaning is different. Therefore it is very important for any lexical ambiguity experiment to use words with frequencies appropriate to the population it is going to test. This is particularly true of balanced ambiguities as not only does the overall frequency impact the experiment (in terms of finding unambiguous frequency matched controls), the relative frequencies of each meaning must also be balanced. We examined several different published works and harvested lexical items from them as a starting point, and then conducted our own on line experiment (which included location and language history data) to identify which of those harvested words were balanced ambiguities for our to-be-tested-population.

### 7.2 Acknowledgement of Sources

We first looked to Duffy et al. (1988) who appended their materials to their paper. Many of these were very America specific (such a *pitcher* and *cell*) and were discarded, but we took thirty items into consideration. We then took a further twenty words from an unpublished study by Roger van Gompel (Personal Communication, 2010). This gave us fifty words, and we generated a further fifty ourselves to bring our total to one hundred.

### 7.3 WebExp Online Experiment

We utilised the British National Corpus (BNC)<sup>2</sup> and a database called SemCor<sup>3</sup> to identify the number of senses each word on the list had, and what they meant. SemCor also provided

---

<sup>2</sup> Data cited herein have been extracted from the British National Corpus Online service, managed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

<sup>3</sup> Part of the WordNet project run by Princeton University (<http://wordnet.princeton.edu>)

frequency counts of each sense of the word which we used to verify the potential of each word to be balanced. Those which were clearly unbalanced were discarded.

Having established a list of 80 potentially balanced ambiguous words we designed an internet experiment to select the most consistently balanced items. We used an XML program known as WebExp 2<sup>4</sup> to verify which items were considered the most balanced among the population from which our Lexical experiments would sample. This consisted of advertising the online experiment solely to students of the University of Edinburgh. A prize draw was used as an incentive to complete the study, which took approximately twenty minutes to complete.

### 7.3.1 Methods

#### *Procedure*

Each of the 80 ambiguous items from the list was presented, in a random order, to the participant with two instructions. Firstly, participants were asked to type the first word that came to mind on observation of the item. This displayed, in the majority of cases, which way the participant leant on the ambiguity. Secondly, participants were instructed to use the item in a sentence. This allowed for the remainder of cases to be categorised (with a few exceptions which were disregarded). The initial word association response was time-logged, from presentation of the stimuli to when the participant started typing their response.

In order to reduce the potential for deception we did not make Nationality a restriction, instead we choose to have participants select native language from a choice of four options (Native British, American English, Other World English, Non-English). Names and email addresses were also collected for the purposes of the prize draw.

Logging only responses that did not exceed a pre-defined threshold of 4000ms, and which could be readily accredited to one of the expected interpretations of the item (by either the word association or sentential response) allowed us to gain insight into the immediate (albeit offline) lexical bias of each individual for each item.

---

<sup>4</sup> Developed by and reported in Keller, Gunasekharan, Mayo, & Corley (2009)

### *Data Analysis*

A 4000ms threshold was assigned and responses were removed if they were initiated after that threshold had been exceeded. Those responses which were initiated after the 4 second threshold were considered to have potentially included higher level cognition, rather than providing an immediate reflection of the interpretation of the ambiguous item.

Responses were tagged with a code for each sense of the word. Where the two components (the immediate word association and the 'use in a sentence' component) aligned a single tag was given, whereas when they did not align the data for that trial was discarded. The results from each participant were collated so that overall percentages could be calculated. These showed how balanced or how biased towards a particular sense each item was.

We found that gaining the required amount of responses to the online experiment happened much more slowly than first anticipated, so we conducted a binomial mixed effects analysis of the data to see whether those that self-identified as native speakers of British English differed in a significant way to those who self-identified as speakers of American or other world English. We did not include those that self-identified as 'non-English'. We also added age as a factor in the analysis. Random slopes were added for both predictors by Items, but not by Subjects because Native-ness and Age remain static for each participant. Data was centred to remove intercollinearity.

After the removal of those who identified as 'non-English', in total we tested 74 participants. Out of that 74, 54 self-identified as Native British, 10 as American, and 10 as speaking 'Other English'. Neither Native-ness nor Age appeared to have any significant influence on the interpretation of the ambiguous words ( $p > 0.1$  in both cases.)

To identify which words had balanced ambiguities within our test population a series of thresholds were established for each word. We defined 20%, 30% and 40% variance thresholds. We established the thresholds by firstly locating a mid point ( $100/\text{number of senses}$ ) and then taking 20%, 30%, and 40% and dividing it by the number of senses we observed a word having (e.g.  $30/4$  for the 30% threshold on a word that had 4 observed senses). We applied this variance around the midpoint so that half was above and half below. In doing so we established Lower and Upper bounds level of variance which scaled

with the number of observed senses of each individual word. Table 7.1 shows an example of how we generated the variance thresholds for a three sense word.

Table 7.1: Method for calculating Variance Thresholds of 20%, 30% and 40% scaling by number of senses. In this case we have 3 senses, as in the three senses of the word *Spirit* (ghost, alcohol, fighting spirit).

100 / # Senses	MidPoint (MP)	Threshold / # Senses	Variance (V)	MP $\pm$ V	Lower Bound	Midpoint	Upper Bound
100 / 3 =	33.3	20 / 3 =	6.6	33.3 $\pm$ 6.6 =	26.7	33.3	39.9
		30 / 3 =	10	33.3 $\pm$ 10 =	23.3		43.3
		40 / 3 =	13.3	33.3 $\pm$ 13.3 =	20		46.6

We then applied the variance thresholds to each word in descending order, starting with 40%. If the percentage observations of each of a word's senses did not fit within the upper and lower bounds it was discarded. If it fitted within the 40% bounds then it was tested against the 30% bounds, and so on. This allowed us to select the most balanced ambiguous words, as reported by our testable population.

For the word *Spirit*, for example, the three senses have the following frequencies: Ghost = 27%, Alcohol = 33%, Fighting Spirit = 40%. As can be seen from Table 7.2 these values fit within the established upper and lower bound of the 40% variance threshold, and the 30% variance threshold, but not the 20% variance threshold.

Table 7.2: An example of the application of variance thresholds to the three sense word *Spirit*.

Spirit	Frequency as %	Within Lower and Upper Bounds for:		
		40% (20-46.6)	30% (23.3 – 43.3)	20% (26.7 – 39.9)
Ghost	27	Yes	Yes	Yes
Alcohol	33	Yes	Yes	Yes
Fighting Spirit	40	Yes	Yes	No

It is worth noting that we disregarded any senses of a word of which the observations accounted for less than 10% of responses, and, using the sentence responses as a guide we conflated any senses that we had pre-defined as different but which turned out to be too similar to reliably distinguish.

### 7.3.2 Results

Table 7.3: Words Selected as most balanced as judged by our testable population.

<b>Word</b>	<b>Sense 1</b>	<b>Sense 2</b>	<b>&lt; Threshold</b>
Appendix	Document Part	Body Part	20%
Bark	Dog	Tree	30%
Cabinet	Wardrobe	Politics	30%
Calves	Baby Cows	Leg Muscles	20%
Change	Alteration	Coins	20%
Chest	Box	Torso	20%
Class	Social Group	Student Level	30%
Coach	Guide	Bus	20%
Fan	Supporter	Air Conditioner	40%
Matter	Physical Stuff	Subject	30%
Mold	Plastic Shape	Fungal Spores	20%
Nails	Woodwork	Body Parts	30%
Panel	Judges	Wood Section	20%
Pupil	Eye	Student	20%
Ruler	Dictator	Measuring Device	20%
Sense	Perception	Logic (Common)	20%
Yard	Measurement	Courtyard	20%

No three or more sense words met the required the 40% or below threshold, so all were discarded. 30 two sense words were selected which fit the 40% or below threshold for being balanced. Out of those 30, 13 were discarded as having senses which were too similar to be reliably differentiated (e.g. Judgement: Decision, Legal Decree). The remaining 17 passed the variance thresholds by the following proportions:

1 was below the 40% threshold, 5 were below the 30% threshold, and 11 were below the 20% threshold. The words selected are displayed in Table 7.3 along with their senses and thresholds.

A final word (Match) was added post-hoc to create a balanced design of 18 items. This word was not tested in the online experiment but was showed to be highly balanced in the corpus analysis (Semcor / BMC) It met the 20% threshold and had two senses: Pairing (as in a football game or partnership, and the strike-able light generating object.

### 7.3.3 Discussion

It was necessary to verify the balanced-ness of the ambiguous words harvested from other sources as the majority of those supplied did not test well against our testable population. It was unfortunate that so many tested poorly as this left us with only 17 words to work with, and some of those did not meet the ideal sub 20% variance threshold.

With regards to the methodology used, we feel that a web experiment is a reasonable way to garner this kind of lexical information and that our experiment was well designed to elicit as near online ambiguity resolution responses as possible given the nature of the platform. It was unfortunate, however, how long this phase of the experiment took, and it should be noted for further research that online tasks which are limited to certain populations need to be heavily advertised and incentivised in order to achieve the desired results in the required time frame.

## 7.4 Unambiguous Controls (British National Corpus)

Having established a list of 18 balanced ambiguous words we utilised the British National Corpus to seek out unambiguous controls which were matched in lexical frequency and word length to our items. Word frequency was matched with the combined frequency of all of the senses of the ambiguous item. Frequency thresholds, similar to those above, were used to identify words that were within 30%, 20%, and 10% of each item, and at this stage word length was allowed to vary by plus or minus one character. Lists were made of those items which matched the above criteria and control items were selected from these lists based on their similarity to the ambiguous item in terms of length, frequency, syntactic class (e.g. noun versus verb) and semantic factors such as animacy. Word length was



prioritised as the intended experiments were going to use eyetracking as the principal methodology and we wanted to keep the critical regions of the sentences as similar as possible. 18 items were selected and are displayed alongside their ambiguous counterparts in Table 7.4 below. Word length was kept the same for each item / control pair, but overall lexical frequency differed: 7 pairs differed by less than 10%, 8 by less than 20%, and 3 by less than 30%.

Table 7.4: Ambiguous items and their Length and Frequency Matched Controls

<b>Ambiguity</b>	<b>Control</b>	<b>Word Length</b>	<b>&lt; Threshold</b>
Appendix	Fragment	=	20%
Bark	Pest	=	30%
Cabinet	Content	=	10%
Calves	Joints	=	20%
Change	Policy	=	10%
Chest	Clock	=	20%
Class	Event	=	20%
Coach	Actor	=	10%
Fan	Cat	=	30%
Match	Radio	=	10%
Matter	Report	=	20%
Mold	Rack	=	10%
Nails	Bolts	=	20%
Panel	Crowd	=	20%
Pupil	Mouth	=	10%
Ruler	Graph	=	10%
Sense	Light	=	20%
Yard	Gate	=	30%

## 7.5 Aim and Style of Final Sentences

We created a target sentence which read as normally as possible with either the ambiguous word or its control as the head noun. An example of this is given below:

It was no good, the appendix would have to be removed. (Ambiguous)  
 It was no good, the fragment would have to be removed. (Control)

Two context sentences were then created: One which was very neutral and contained very little propositional content, and one which supported the subordinate sense of the ambiguous word. In the example given below, our population (of students) reported that the subordinate sense of Appendix was the body part, thus the contextually supportive sentence references the abdomen:

The woman was unhappy with the way things were going. (Neutral)  
 The woman was complaining of abdominal pain. (Supportive)

We crossed these two factors to make a 2x2 manipulating Context and Lexical Ambiguity. An example of this can be seen in 7.5.a- 7.5.d below, and a full list of materials can be found in Appendix 2.

7.5.a. The woman was unhappy with the way things were going.  
 It was no good, the fragment would have to be removed.  
*(Neutral Unambiguous)*

7.5.b. The woman was unhappy with the way things were going.  
 It was no good, the appendix would have to be removed.  
*(Neutral Ambiguous)*

7.5.c. The woman was complaining of abdominal pain.  
 It was no good, the fragment would have to be removed.  
*(Supportive Unambiguous)*

7.5.d. The woman was complaining of abdominal pain.  
 It was no good, the appendix would have to be removed.  
*(Supportive Ambiguous)*

## **Chapter 8: NoLoad, Context, and Lexical Ambiguity (Experiment 3)**

### **8.1 Introduction**

This is the first in a series of experiments investigating how the availability of cognitive resources influences the role of context on the processing of Lexical ambiguities. The structure of the series is similar to that reported in the Syntax section; first we will present a non-loaded baseline experiment, then a second experiment which is the same in all respects except that the participants are placed under a cognitive load (digit span – see Section 1: Chapter 2), and then we will present a statistical meta-analysis of the two. Differing from that previous structure we will also report an additional experiment in which the load manipulation was within subjects rather than between.

In the introduction to this section we have discussed some of the literature regarding the role of context and resources with regard to lexical ambiguity resolution; it became apparent that there are two main types of lexical ambiguity which respond differently to contextual support (Balanced and Biased ambiguities (or in the rather unintuitive terminology of Rayner & Duffy, 1986, equibaised and non-equibaised)). Duffy et al. (1988) suggested a model called the Reordered access model which describes the mechanisms by which these two types of ambiguity are influenced by a preceding context, or one presented afterwards which disambiguates the ambiguity downstream. Miyake et al. (1994) then investigated how working memory capacity interacted with biased ambiguities which are disambiguated by a context downstream. As has been discussed above, we will be using Balanced ambiguities and a preceding context (creating a 2x2 design of Context x Ambiguity). This is partly because this area has not been investigated (to our knowledge), and partly because the predictions made by Duffy et al.'s Reordered Access model and Miyake et al.'s Capacity Constrained model of lexical ambiguity resolution are more clearly testable for balanced ambiguities than biased ambiguities, in the case of a preceding context. In the previous chapter we have discussed the generation of the materials which we will be using.

As in the syntax section, we will be using fully online eyetracking to gain a comprehensive picture of eye movement behaviour during reading (including regressions and re-readings which the moving window paradigm used by Miyake et al. (1994) and others does not

provide). Also, for the second and fourth experiments in the series, rather than relying on the pre-definition of those participants with categorically low, medium, and high reading spans, we will experimentally manipulate the resources available by titrating the difficulty of the digit span task such that each participant has a similar amount of available resources.

The main predictions of this non-resource limited experiment are:

A main effect of Ambiguity on or just after the ambiguous word. This represents the increased cost of activating two meanings and selecting between them which is associated with an ambiguity as compared to an unambiguous control.

A main effect of Context across all regions such that a supportive context speeds processing. This represents a corollary to the discourse cohesion effect observed in the syntax section: Readers find it easier to process text within a context than in isolation, regardless of particular challenges such as ambiguity.

An interaction of Context and Ambiguity on and in the regions following the ambiguous word such that when a supportive context is present there is no (or very limited) effect of ambiguity, but when it is absent a larger effect of ambiguity occurs. This interaction represents contextual facilitation of selection, versus an increase of processing load associated with maintaining two interpretations until the end of the sentence (in the hope of locating some disambiguating information downstream).

## 8.2 Methods

### *Participants*

36 participants meeting the following experimental criteria took part in the study.

All were Native speakers of British English, over 18 years of age, had no known reading disorders (e.g. dyslexia), and had normal or corrected to normal vision.

Total Exclusions based on comprehension questions and poor tracking = 3

Comprehension Question Range (of included participants) = 85% - 98%, Mean = 92%.

### *Stimuli*

The 36 experimental stimuli were created as detailed in Section 3: Chapter 7 (See Appendix 2 for full list). Each item began with a sentence which either provided no supportive context at all, or supported the non-dominant interpretation of the upcoming critical noun (Neutral / Supportive). After the context sentence, a brief lead in preceded the critical noun and a brief lead out allowed for spillover and wrapup regions. The critical noun was either a frequency and length matched control, or a lexically ambiguous item where the two or more meanings were of equal lexical frequency (Unambiguous / Ambiguous). A set of example stimuli is given in 8.2.a - 8.2.d

8.2.a. The woman was unhappy with the way things were going.  
It was no good, the fragment would have to be removed.  
(Neutral Unambiguous)

8.2.b. The woman was unhappy with the way things were going.  
It was no good, the appendix would have to be removed.  
(Neutral Ambiguous)

8.2.c. The woman was complaining of abdominal pain.  
It was no good, the fragment would have to be removed.  
(Supportive Unambiguous)

8.2.d. The woman was complaining of abdominal pain.  
It was no good, the appendix would have to be removed.  
(Supportive Ambiguous)

### *Procedure*

The 144 experimental materials (36 items in 4 conditions each) were divided into 4 lists (Latin Square design). The experimental items, along with 72 fillers, were presented in a pseudo randomised order such that each participant saw only one condition of each item. The randomisation disallowed two experimental items to occur consecutively and made sure that a filler always intervened. All equipment and general procedures were unchanged from those described in the Syntax Section (Section 2: Chapter 3).

## Data Analysis

The experimental items were broken down into six regions. The first region comprised the context sentence (Context region). The second (non-context) sentence was broken down into five regions. The first included all words up to the word preceding the unambiguous / ambiguous noun (Initial region). The second was the word preceding the unambiguous / ambiguous noun (Pre-critical region). The third was the unambiguous / ambiguous noun (Critical region). The fourth region was the ten or so characters following the critical region, rounded to the nearest word (Spillover region). The fifth was the rest of the sentence (Wrapup region). A further region was created by combining the Pre-Critical and Critical regions. An example of the region breakdown is given in 8.2.e:

### 8.2.e. Region Breakdown

/The woman was complaining of abdominal pain./

/ 1 /

/It was no good,/ the /appendix /would have/ to be removed./

/ 2 / 3 / 4 / 5 / 6 /

1 = Context, 2 = Initial, 3 = Pre-Critical, 4 = Critical, 5 = Spillover, 6 = Wrapup

The same fixation duration (*First Pass*, *Second Pass*, *Total Time*, *Regression Path*, and *First Fixation*) and eye movement control measures (*Fixation Probability*, *Regressions In*, and *Regressions Out*) were used as in Section 2: Chapter 3, and the data analysis procedure described in that section was also maintained. We added an additional predictor to the analysis labelled Frequency, which was a continuous predictor designed to check that no undue influence was being contributed by the frequency of the ambiguity / frequency matched control. Frequency is often confounded with word length so we used a linear model to model the collinearity of the two factors and extract the residuals. Linear mixed effects models were run including the continuous residual Frequency factor, but for simplicity we report residual Frequency as a categorical two level factor (High and Low) generated by a median split along the centred value.

## 8.3 Results

*/It was no good /the /appendix /would have/ to be removed./*  
*/        2        / 3 /        4        /        5        /        6        /*

1 = Context, 2 = Initial, 3 = Pre-Critical, 4 = Critical, 5 = Spillover, 6 = Wrapup

### Overview of Critical Findings:

In this NoLoad experiment we do not observe a Context by Ambiguity interaction in the critical region in first pass (as did Duffy et al., 1988), but do see a Context by Ambiguity interaction on the probability of fixating the critical word (see Figure 8.1). Readers are more likely to fixate the critical word if it is ambiguous, and the context does not support a particular interpretation, but if the context is supportive no difference in the likelihood of fixation is observed due to ambiguity. This implies that readers are able to utilise context to help with lexical disambiguation prior to fixating the ambiguous word, and that they can adapt their fixation behaviour accordingly. This interaction is supported by a Context by Ambiguity interaction in the Pre-Critical region which patterns slightly differently (see Figure 8.2).

### Context by Ambiguity Interactions:

#### Probability of Fixating the Critical Region:

We observe an interaction between Context and Ambiguity in the likelihood that the Critical Region is fixated ( $p = 0.009^{**}$ ) (see Figure 8.1). There are marginal pairwise effects across Context both for Unambiguous and Ambiguous critical words. For Unambiguous Critical Words we observe a higher probability of fixation if the Context is Supportive (Neutral = 59% vs Supportive = 66%,  $p = 0.052^m$ ), and for Ambiguous Critical Words we see the reverse pattern (Neutral = 70% vs Supportive = 64%,  $p = 0.069^m$ ). When the Context is Neutral, we see a significantly greater likelihood of fixation for Ambiguous words, as compared to their Unambiguous controls (70% vs 59%,  $p = 0.001^{**}$ ). When the Context is Supportive there is no difference in the probability of fixation for Ambiguous and

Unambiguous words (66% vs 64%,  $p > 0.1$ ). This implies, not only that people are sensitive to Ambiguity when the Context does not support one interpretation, but that their eye movements can be guided by ambiguity prior to actual fixation on the ambiguous word. This has important implications for Eye Movement Control models such as EZ-Reader (Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 1999; Reichle, Warren, & McConnell, 2009) and Swift (Engbert, Longtin, & Kliegl, 2002; Engbert & Nuthmann, 2005; Nuthmann & Engbert, 2009; Richter, Engbert, & Kliegl, 2006) as it indicates that certain lexical properties may be processed in the para-fovea (Balota, Pollatsek, & Rayner, 1985; Inhoff & Rayner, 1986; Rayner, Inhoff, Morrison, Slowiaczek, & Bertera, 1981), or via a predictive mechanism. The precise mechanics of computation models of eye-movement control are not within the scope of this thesis, but we will discuss the implications of this in more detail in section 8.4.

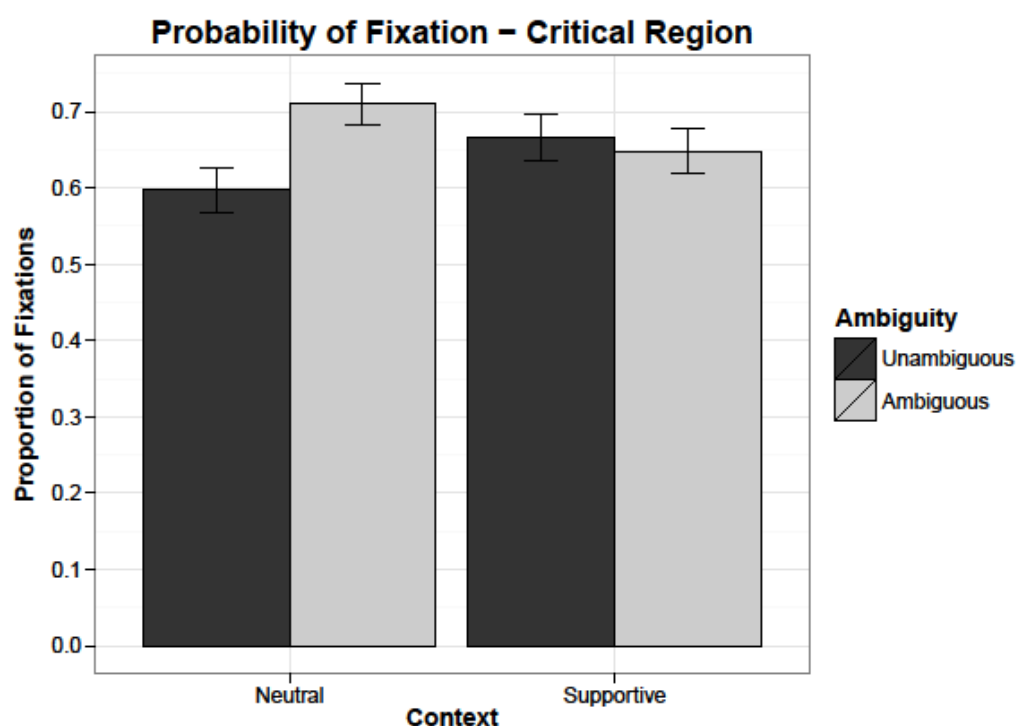


Figure 8.1: Experiment 3 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context x Ambiguity Interaction.

We also observe a Context by Ambiguity interaction in the two previous regions (Initial and Pre-Critical). We are not going to discuss the interaction in the Initial Region as we only



have 72 observations of the region being skipped (compared to 1222 observations of fixation) and this makes drawing any conclusions about the behaviour of individual factors slightly dubious.

#### Probability of Fixating the Pre-Critical Region:

The interaction in the Pre-Critical region ( $p = 0.034 *$ ) is of interest, however, and supports the assertion above that some level of para-foveal or predictive processing is occurring and guiding eye-movement control.

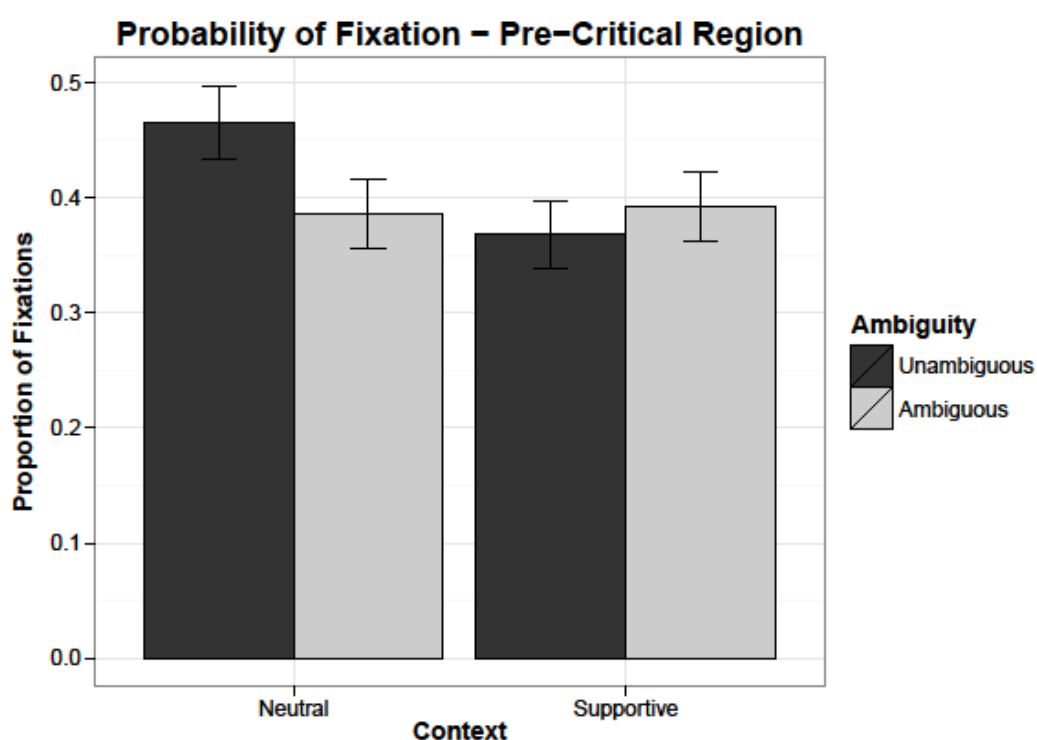


Figure 8.2: Experiment 3 - Proportion of Fixations for Probability of Fixation on the Pre-Critical Region. Error Bars display Standard Error. Context x Ambiguity Interaction.

We see a similarly flat pattern when the Context is Supportive (Ambiguous = 39% vs Unambiguous = 36%,  $p > 0.1$ ), but when the Context is Neutral we see the reverse of what we observed in the Critical Region (Ambiguous = 38% vs Unambiguous = 46%,  $p = 0.024 *$ ) (see Figure 8.2). We also see that there is a significant pairwise difference across Context when the upcoming Critical Word is Unambiguous (Neutral = 46% vs Supportive = 36%,  $p = 0.005 **$ ) but no such effect when the Critical Word is Ambiguous (Neutral = 38% vs

Supportive = 39%,  $p > 0.1$ ) We interpret this as an attraction effect: Upon reaching the end of the Initial Region and planning a saccade, the Pre-Critical Region is skipped over in favour of fixating the upcoming Critical Region more often if the Critical Region is Ambiguous, than if it is Unambiguous but only if the Context is Neutral and hasn't allowed for any prediction or provided any support to para-foveal processing.

### **Other Interactions:**

In the sections below we report interactions involving Frequency. These are not central to our investigation but indicate that our participants are sensitive to low level factors like lexical frequency.

To anticipate our results, we see the following interactions:

Ambiguity by Frequency interactions:

#### *First Fixation on the Spillover Region:*

Unambiguous words appear slightly more influenced by Frequency than do Ambiguous words. When unambiguous, high frequency words receive shorter fixations than low frequency words, but this pattern, with a smaller difference, is reversed for ambiguous words (see Figure 8.3).

#### *Second Pass for the Spillover Region:*

In second pass the reverse pattern of the First Pass finding occurs. When unambiguous, high frequency words receive longer fixations than low frequency words, but this pattern, is reversed for ambiguous words (see Figure 8.4).

#### *Regression Path of the Wrapup Region:*

In regression path of the wrapup region, we see a similar pattern to first pass spillover for unambiguous words (shorter reading times for high frequency words), and this pattern is replicated, but with a smaller (and non-significant) difference for ambiguous words (see Figure 8.5).

### Ambiguity by Frequency Interaction:

#### First Fixation on the Spillover Region:

For First Pass we observe an interaction between Frequency and Ambiguity ( $p = 0.038 *$ ) in the Spillover Region in which we see longer fixation durations for Low Frequency words compared to High Frequency words when the critical word is Unambiguous (249ms vs 232ms,  $p > 0.1$ ), and longer fixation durations for High Frequency words compared to Low Frequency words when the critical word is Ambiguous (252ms vs 240ms,  $p > .05$ ) (see Figure 8.3). These pairwise comparisons are not significant. This interaction is not particularly interesting in terms of the predictions of the experiment as it does not involve context. However, it is reported here to demonstrate that our participants are sensitive to low level factors like lexical frequency.

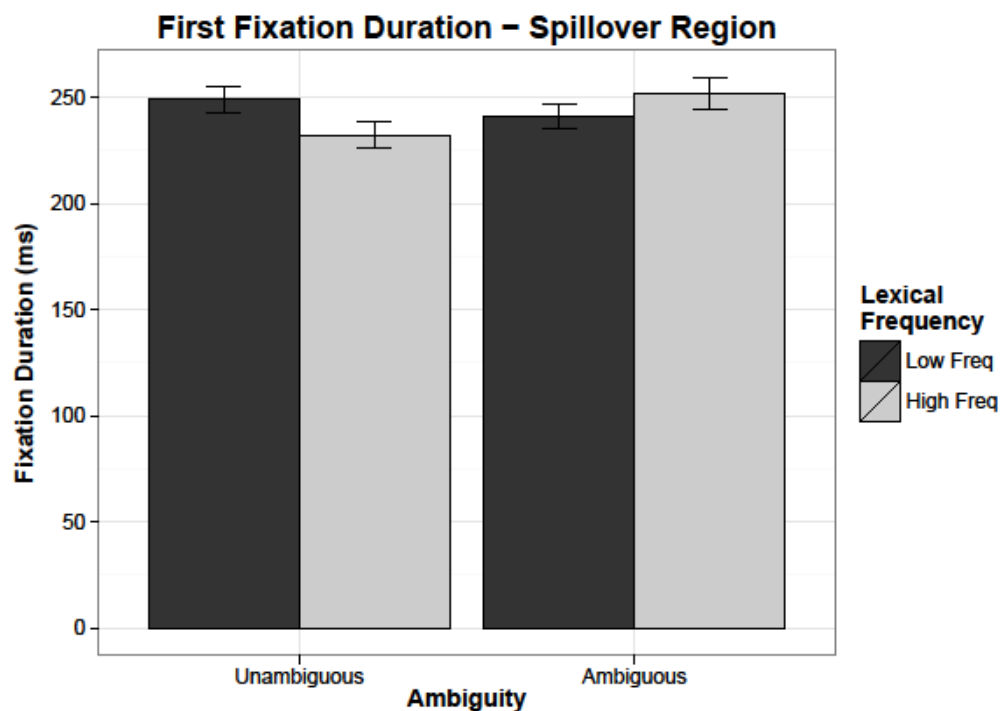


Figure 8.3: Experiment 3 - Mean fixation durations for First Fixation in the Spillover Region. Error Bars display Standard Error. Ambiguity x Lexical Frequency Interaction.

### *Second Pass for the Spillover Region:*

In Second Pass we see an Ambiguity by Lexical Frequency interaction in the Spillover Region ( $p = 0.034^*$ ). We observe the reverse pattern that we saw in First Fixation, implying that there is some later processing occurring which counters the initial effect. Here we see longer fixation durations for High Frequency words compared to Low Frequency words when the word is Unambiguous (237ms vs 202ms,  $p > 0.1$ ), as opposed to longer fixation durations for Low Frequency words compared to High Frequency words when the word is Ambiguous (225ms vs 185ms,  $p = 0.073^m$ ) (see Figure 8.4).

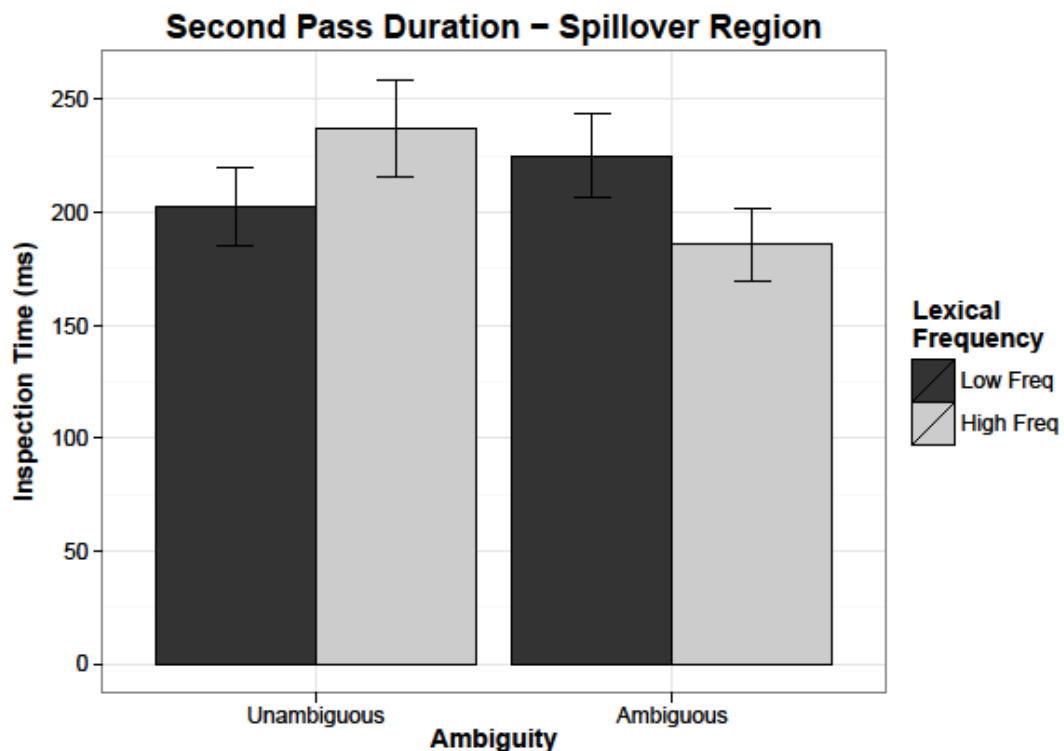


Figure 8.4: Experiment 3 - Mean inspection time for Second Pass in the Spillover Region. Error Bars display Standard Error. Ambiguity x Lexical Frequency Interaction.

### *Regression Path of the Wrapup Region:*

In the Regression Path of the Wrapup Region we see another Ambiguity by Frequency interaction ( $p = 0.019^*$ ). Compared to the observed effects in the Spillover Region, we see a similar pattern for Ambiguous words (i.e. longer fixation durations if the word is Low

Frequency compared to High Frequency (137ms vs 124ms,  $p > 0.1$ ), but the reverse pattern for Unambiguous words (i.e. longer fixation durations for Low Frequency words compared to High Frequency words (156ms vs 101ms,  $p = 0.039^*$ )). This leaves us with the same direction for both Unambiguous and Ambiguous words (longer fixations durations if the word is Low Frequency) but the magnitude of the difference is far greater for Unambiguous words (see Figure 8.5). Again, these interactions are not central to our discussion but support the observation that our participants are sensitive to lower level factors. It is interesting that we do not observe the higher level context factor interacting with them in the reading measures.

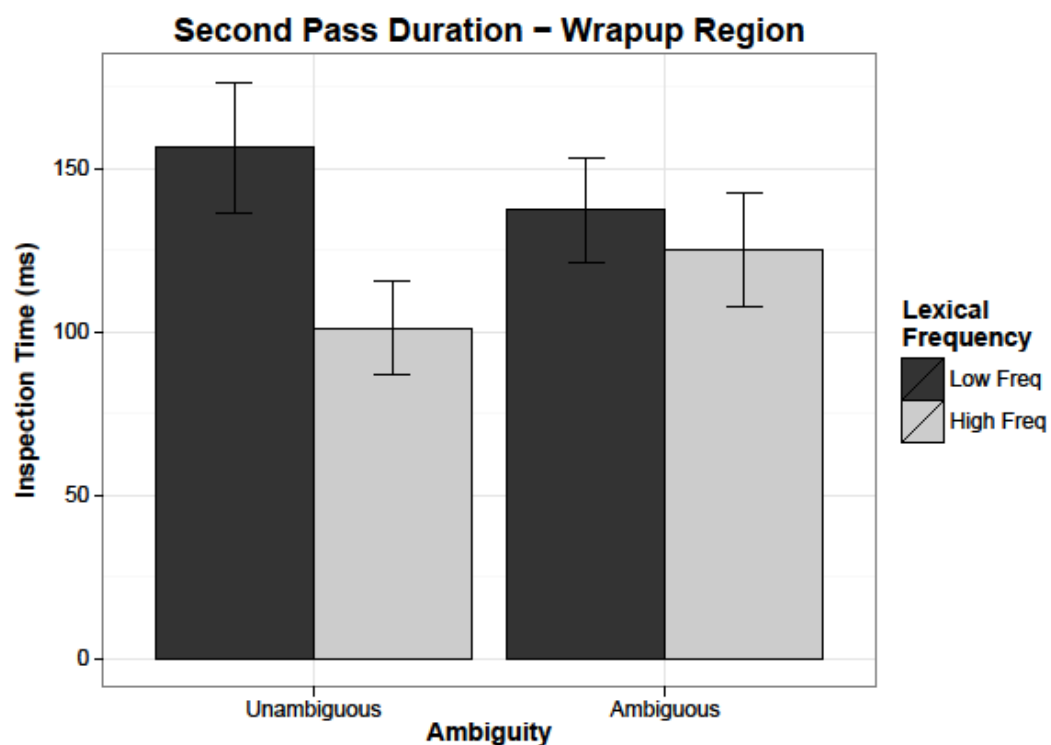


Figure 8.5: Experiment 3 - Mean inspection time for Second Pass in the Wrapup Region. Error Bars display Standard Error. Ambiguity x Lexical Frequency Interaction.

### **Summary of Main Effects:**

The main effects are reported in the tables below, Table 8.1 reports Context effects, while Table 8.2 reports Ambiguity effects.

#### **Context:**

As can be seen from Table 8.1, Context appears to have little effect on the early reading measures, instead coming in to effect more in re-readings (SP, TT) and regressions (RI). In all cases having a supportive context facilitates processing.

#### **Ambiguity:**

There is very little influence of ambiguity on reading measures or regressions in this experiment (see Table 8.2), but it does appear to influence the probability of fixating the critical region. The likelihood of fixating the critical region is increased if the noun is ambiguous (this main effect is also part of a Context by Ambiguity interaction reported above), but interestingly, when the noun is ambiguous, the likelihood of fixating the subsequent Spillover region is reduced. Presumably this is due to increased processing in the para-fovea for those fixating the noun.

#### **Frequency:**

We do not provide a table for frequency as it is not of principal interest to the experiment, but we should mention a main effect of frequency in the Pre-Critical region such that the frequency of the upcoming noun impacts the first pass reading time of the preceding word. Longer reading times are observed when the upcoming noun is of low frequency, rather than high frequency. This is likely due to pre-processing in the para-fovea.

Table 8.1: Experiment 3 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if Context is Neutral or Supportive. Effects are also marked if Marginal.

<b>Context</b>	<b>Levels</b>	<b>Initial</b>		<b>Pre-Critical</b>		<b>Critical</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Pre-Crit + Crit</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Neutral	1.58	754.1	-0.07	231.0	-0.78	240.4	0.38	353.3	-1.50	390.0	-1.11	299.1
	Supportive		787.6		234.7		233.3		353.8		365.2		268.1
First Fixation	Neutral	-1.27	221.4	0.52	220.1	-0.83	227.9	0.54	242.2	-0.58	240.2	-0.17	229.1
	Supportive		217.0		225.7		223.0		244.8		238.5		227.5
Regression Path	Neutral	1.30	799.9	-0.55	310.9	-0.49	331.1	-1.77 m	713.6	-2.07 *	1349	-0.82	392.3
	Supportive		824.6		306.4		330.9		662.9		1242		382.9
Second Pass	Neutral	-1.47	300.1	-2.40 *	89.06	-1.02	116.6	-2.02 *	233.1	-1.52	141.7	-1.63	194.8
	Supportive		273.0		68.47		106.6		192.1		118.1		168.2
Total Time	Neutral	-0.50	1014	-2.16 *	305.5	-0.74	340.5	-2.55 **	569.5	-2.69 **	566.4	-3.32 ***	477.0
	Supportive		1013		284.5		334.5		527.4		510.3		434.9
Probability of Fixation	Neutral	-0.74	.95	-1.78 m	.41	0.16	.66	-1.83 m	.90	0.85	.81	-1.23	.89
	Supportive		.94		.37		.66		.80		.83		.86
Regressions Out	Neutral	-0.18	.02	0.43	.18	-0.80	.21	0.00	.33	-1.26	.75	0.26	.16
	Supportive		.02		.17		.21		.30		.73		.17
Regressions In	Neutral	-1.88 m	.49	-2.26 *	.31	-2.62 *	.37	-0.74	.38	/	/	-3.07 **	.49
	Supportive		.45		.25		.31		.36		/		.41

Table 8.2: Experiment 3 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if the Critical Noun is Ambiguous or Unambiguous. Effects are also marked if Marginal.

<b>Ambiguity</b>	<b>Levels</b>	<b>Initial</b>		<b>Pre-Critical</b>		<b>Critical</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Pre-Crit + Crit</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unambiguous	-0.49	766.6	-0.76	239.2	-0.56	237.2	1.15	343.1	-0.05	378.9	-0.81	29.7
	Ambiguous		775.1		226.5		236.5		364.1		376.3		288.5
First Fixation	Unambiguous	-0.41	220.5	-0.65	227.5	-0.80	226.1	0.70	240.6	0.07	237.9	-0.66	228.7
	Ambiguous		217.9		218.3		224.8		246.4		240.7		227.9
Regression Path	Unambiguous	-1.21	817.4	-0.33	309.4	-0.43	337.0	-0.32	699.6	-0.41	1321	-0.45	392.4
	Ambiguous		807.1		307.9		325.0		676.9		1270		382.9
Second Pass	Unambiguous	-1.26	300.2	-0.65	81.79	-0.30	113.8	-0.85	219.8	-0.20	128.7	-0.58	185.3
	Ambiguous		272.9		75.73		109.4		205.5		131.1		177.7
Total Time	Unambiguous	-1.17	1019	-1.06	304.0	-0.04	340.1	-0.59	549.0	-0.02	541.3	-0.37	462.4
	Ambiguous		1008		286.0		334.9		547.9		535.4		449.5
Probability of Fixation	Unambiguous	1.28	.94	-1.22	.41	2.03	.63	-2.01	.90	0.68	.82	1.10	.86
	Ambiguous		.95		.38	*	.69	*	.80		.83		.89
Regressions Out	Unambiguous	-0.18	.03	0.43	.17	-0.80	.22	0.00	.31	-1.26	.75	0.12	.17
	Ambiguous		.03		.18		.19		.32		.73		.17
Regressions In	Unambiguous	-0.52	.47	0.06	.28	0.26	.34	0.16	.37	/	/	0.47	.41
	Ambiguous		.48		.28		.35		.38		/		.46



## 8.4 Discussion

We set out to replicate Duffy et al. (1988)'s findings regarding balanced ambiguous words and preceding contexts as a way of establishing a no-Load base line to which we could compare participants under conditions of limited resources.

Duffy et al. showed a Context by Ambiguity interaction on the ambiguous word such that longer reading times were observed when the disambiguating context did not precede the ambiguity, compared to when it did. We did not observe this interaction, and so it is hard to verify the predictions of the reordered access model. There are two reasons that we may not have observed this same interaction, however, neither of these are particularly satisfactory. Firstly, our supportive context was similar to Duffy et al.'s *before* context, but our Neutral context was visually very different to their *after* context. In Duffy's *after* context less text preceded the ambiguity than in our neutral context, and more came after it. It is possible that in the Duffy experiments, because readers were able to see that lots of information came after the ambiguity, they may have committed to adopting and elaborating two interpretations in anticipation of some disambiguating information downstream. In our experiments however, the ambiguity was comparatively close to the end of the text, and had been preceded by a very neutral context; it is possible that readers chose to commit to one interpretation on the spot (driven purely by frequency) as it was clear that maintaining two would not be profitable as no disambiguating information was likely to appear. This explanation is shaky however, as we would have expected to see a main effect of ambiguity in the critical region if readers were engaging in selection. The second explanation is that Duffy et al. used a measure called Gaze Duration instead of the First Pass that we report here. Gaze duration expanded the critical region to include the final 4 characters of the previous region, if no fixation was made on the normal critical region. First Pass does not include the previous 4 characters if the region is not fixated, but this methodological explanation is also unsatisfactory because we replicated the Gaze Duration measure and still did not find the same effect.

A third explanation is driven by our observation of a Probability of Fixation interaction for the critical region such that readers were more likely, in a neutral context, to fixate the critical word if it was ambiguous than if it was not, whereas in a supportive context there was no difference. Given that we see selective fixation of the critical region based on the

propositional content of the context and the presence of an ambiguity, it isn't that surprising that we do not see these factors interact on the first pass of the region itself.

One strong interpretation of this finding is that participants are able to begin undertaking lexical access and the generation of meanings during para-foveal processing and not only recognise an ambiguity, but also determine whether or not the context has provided enough support for them to be able to skip it or not. For this to guide eye movements it would have to happen quickly enough for the reader to have time to program a saccade to, or over the word. To couch our discussion in terms of the EZ-Reader model, this is not possible, both because lexical access (and the activation of meanings) to occur at the second stage of processing ( $L^2$ ) which can only occur under direct fixation, and the time taken to program (and possibly cancel and reprogram) a saccade is greater than the average time spent on a function word such as 'the' (which is where the fixation would need to be for para-foveal processing to be taking place) (Reichle et al., 2009). It may be possible within the SWIFT framework as that allows for parallel processing of multiple words at once such that the processing of  $n$  and  $n+1$  can occur at the same time, and even finish in a non-linear order (Richter et al., 2006). A more reasonable interpretation might be that readers are able on some level to recognise the presence of an upcoming ambiguity (while not actually performing lexical access to retrieve meanings) and check to see if the context has been particularly full of lacking in propositional content that might help. If it has not they may decide more often that they need process the word fully under direct fixation, but if it has they may feel that they can cope without doing so. This is also not possible under the EZ-Reader architecture as, while  $L^1$  processing (a familiarity check) can occur in the para-fovea, the timing for programming a saccade remains problematic, and information about if a word is ambiguous or not is the purview of  $L^2$  processing rather than  $L^1$ . A third possibility is that the upcoming ambiguous word is predicted by the context well enough that participants are able to skip the fixation entirely in the case of a supportive context; however, while this would be acceptable from the perspective of eye movement control models, it wouldn't explain the disparity between likelihood of fixation for unambiguous and ambiguous words in the neutral context condition.

While it is interesting to consider the origins of this effect, for the purposes of this thesis we are not particularly concerned with the precise functioning of the mechanism behind it,

only that it occurs and indicates that we can detect contextual facilitation of lexical ambiguity processing on some level in our lab.

In the next chapter we report a replication of this experiment with participants under a cognitive load but, given that in this experiment we did not observe the Context by Ambiguity interaction that Duffy et al. showed, we cannot predict how our results will sit with the predictions of Miyake et al.'s capacity constrained model (probability of fixation is not discussed within the model).

## **Chapter 9: Load, Context, and Lexical Ambiguity (Experiment 4)**

### **9.1 Introduction**

In the previous experiment we established that readers were in general sensitive to the presence of a context which was full of propositional content, rather than a neutral one. This manifested mostly in the later reading measures and regressions in to the critical regions.

We did not see much influence of the ambiguity itself in any reading measures but there was a main effect of ambiguity in the probability of fixating the critical region such that more fixations were made if the critical word was ambiguous.

Importantly we did not observe the anticipated Context by Ambiguity interaction in the critical region (as was reported by Duffy et al., 1988), but instead saw that readers fixated on the critical region (containing the ambiguous word or frequency and length matched control word) more if the word was ambiguous and the context was not supportive of a particular interpretation.

In this experiment we will apply a titrated extrinsic memory load to experimentally reduce the availability of cognitive resources. The procedure for this is the same as it was in the previous Syntax experiments (Section 2). We cannot make specific predictions based on Miyake et al.'s capacity constrained model as the results of our No-Load experiment did not fit within the scope of the model. Taking a step back however, the single resource theory (SR – see Literature Review, section 1.3.1) would predict that participants will not be able to make use of the context as much under load, and so the main effects of context will be reduced or eradicated. The SLIR, on the other hand, would predict that normal contextual facilitation of processing will continue unhindered by load as it is in the interpretive system which draws from a separate resource pool. With regards to the Context by Ambiguity interaction in the probability of fixating the critical region, both SR and SLIR theory would predict that this interaction will not be observed as, in the case of SR: with reduced resources, modularity between language processes increases, and in the case of SLIR: Using context to aid in Lexical disambiguation is a post-interpretive process which is influenced by a reduction in general resources.

## 9.2 Methods

### *Participants*

36 participants took part in the study. The requirements were the same as the no load experiment, with the addition that participants had not taken part in that experiment.

Total Exclusions based on poor comprehension, tracking, or digit recall = 5

Comprehension Question Range (of included participants) = 85% - 98%, Mean = 92%.

Digit Span Range (of included participants) = 3 – 9, Mean = 6.

### *Stimuli*

In all ways, the sentential stimuli were identical to those in experiment three. The presentation of the digits was the same as in the loaded syntax experiment (experiment two).

### *Procedure*

In all ways the procedure was identical to that in experiment 3 with the exception of the digit span pre-test, and subsequently differing trial presentation. This pre-test procedure and subsequent trial presentation were the same as the in the Loaded syntax (experiment two) and have been described in Section 2: Chapter 4.

### *Data Analysis*

In order to maintain comparability, the same procedures were used to analyse this experiment, as were used to analyse experiment three.

### 9.3 Results

*/It was no good /the /appendix /would have/ to be removed./*  
*/        2        / 3 /        4        /        5        /        6        /*

1 = Context, 2 = Initial, 3 = Pre-Critical, 4 = Critical, 5 = Spillover, 6 = Wrapup

#### Overview of Critical Findings:

In the NoLoad experiment we saw a Context by Ambiguity interaction on the probability of fixating the critical word. Readers were more likely to fixate the critical word if it was ambiguous, and the context did not support a particular interpretation. Under Load this interaction is not observed. Instead we see a marginal main effect of ambiguity such that ambiguous words are more likely to be fixated than non-ambiguous words, regardless of context. The implication of this is that readers under load are unable to use context as effectively to (partially) disambiguate the ambiguity in advance. Muddying the waters, however, is a Context, Ambiguity and Frequency interaction in which low frequency words are not affected by an interaction between Context and Ambiguity, but high frequency words are. The pattern for high frequency words is not the same as for the NoLoad experiment however, so it is not the case that being under Load impacts only low frequency words. See Figure 9.1 for more details.

Two other Context by Ambiguity interactions are observed. One in the WrapUp region for First Pass (see Figure 9.2), and in the Pre-Critical region for Total Time (marginal, see Figure 9.3). Neither was present in the NoLoad experiment, and neither is particularly theoretically interesting. Nevertheless we report them here as they involve our principal manipulations.

### Context by Ambiguity Interactions:

#### Probability of Fixating the Critical Region:

We observe a three way interaction between Context, Ambiguity, and Lexical Frequency in the likelihood of fixating the Critical Region ( $p = 0.047 *$ ) (see Figure 9.1 & Table 9.1). When the Critical Word is Low Frequency we see very little variation across conditions (Neutral / Unambiguous = 58% vs Neutral / Ambiguous = 61%,  $p > 0.1$ ) (Supportive / Unambiguous = 62% vs Supportive / Ambiguous = 58%,  $p > 0.1$ ). When the Critical Word is High Frequency, however, we see equally little when the Context is Neutral (Unambiguous = 56% vs Ambiguous = 57%,  $p > 0.1$ ) but when the Context is Supportive we see a highly significant difference between Unambiguous and Ambiguous conditions (45% vs 70%,  $p < 0.001 ***$ ). This interaction is significant ( $p = 0.007 **$ ). In addition, we see a significant difference between the Neutral / Ambiguous and Supportive / Ambiguous conditions when the Critical word is High Frequency (57% vs 70%,  $p = 0.034 *$ ). It appears that when the Critical Word is of High Frequency, Ambiguity has little influence on the likelihood of fixation if the Context is Neutral. If the Context is Supportive however, Ambiguous words are more likely to be fixated than Unambiguous words. This will be discussed in greater detail in this chapter's discussion section (9.4).

Table 9.1: Experiment 4 - Conditional Proportion of Fixations in % for Probability of Fixation: Critical Region. Context, Ambiguity, & Lexical Frequency Interaction.

Context	Lexical Frequency	Low Frequency	High Frequency
Neutral	Unambiguous	58	56
Neutral	Ambiguous	61	57
Supportive	Unambiguous	62	45
Supportive	Ambiguous	58	70

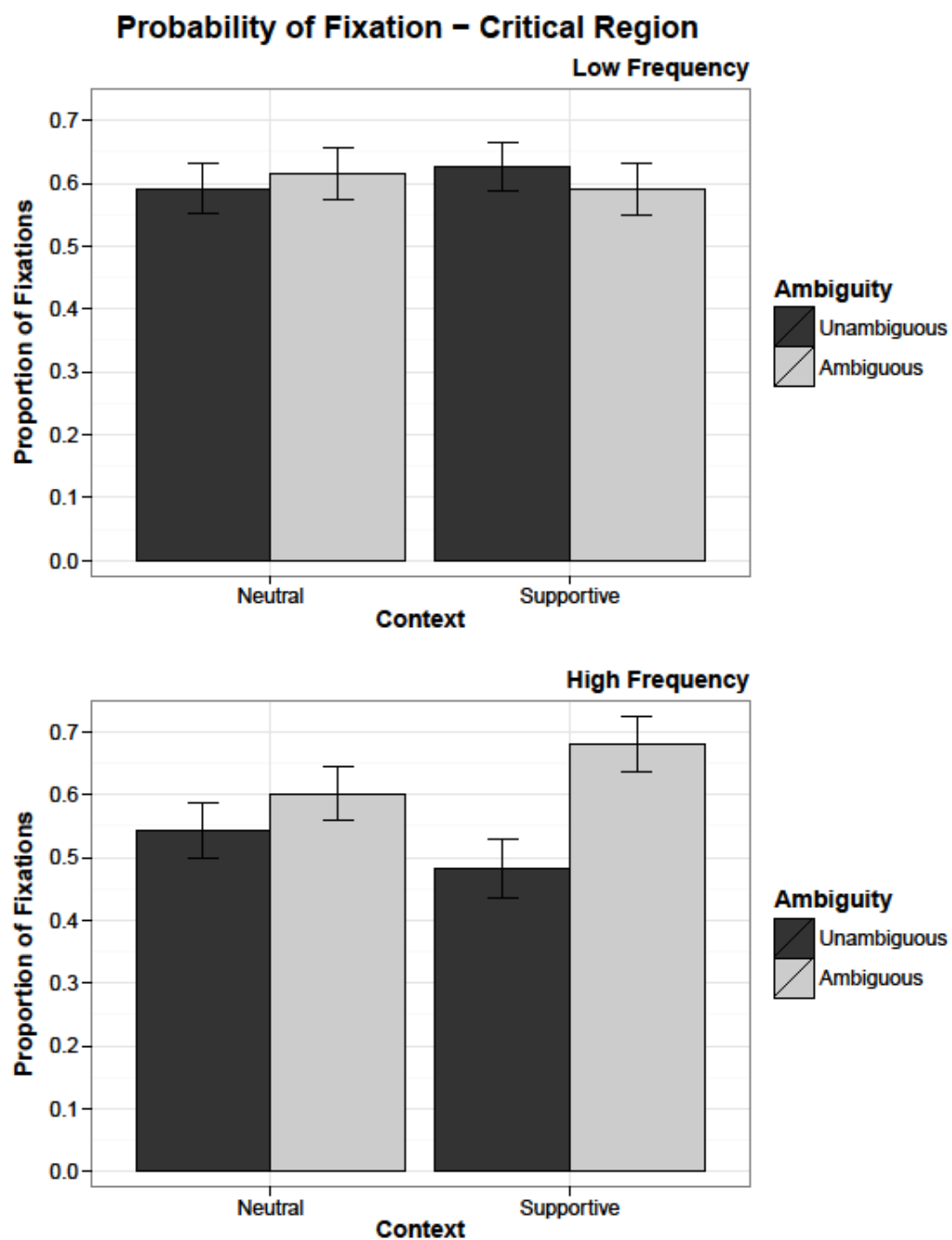


Figure 9.1: Experiment 4 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context, Ambiguity, and Lexical Frequency Interaction.



### First Pass of the WrapUp Region:

In the Wrapup regions we see a Context by Ambiguity interaction ( $p = 0.028 *$ ) (see Figure 9.2). When the Context is Neutral we see no difference between Unambiguous and Ambiguous words (419ms vs 426ms,  $p > 0.1$ ) but when the Context is Supportive we observe a clear slow down for the Unambiguous condition (Unambiguous = 482ms vs Ambiguous = 404ms,  $p = 0.011 *$ ).

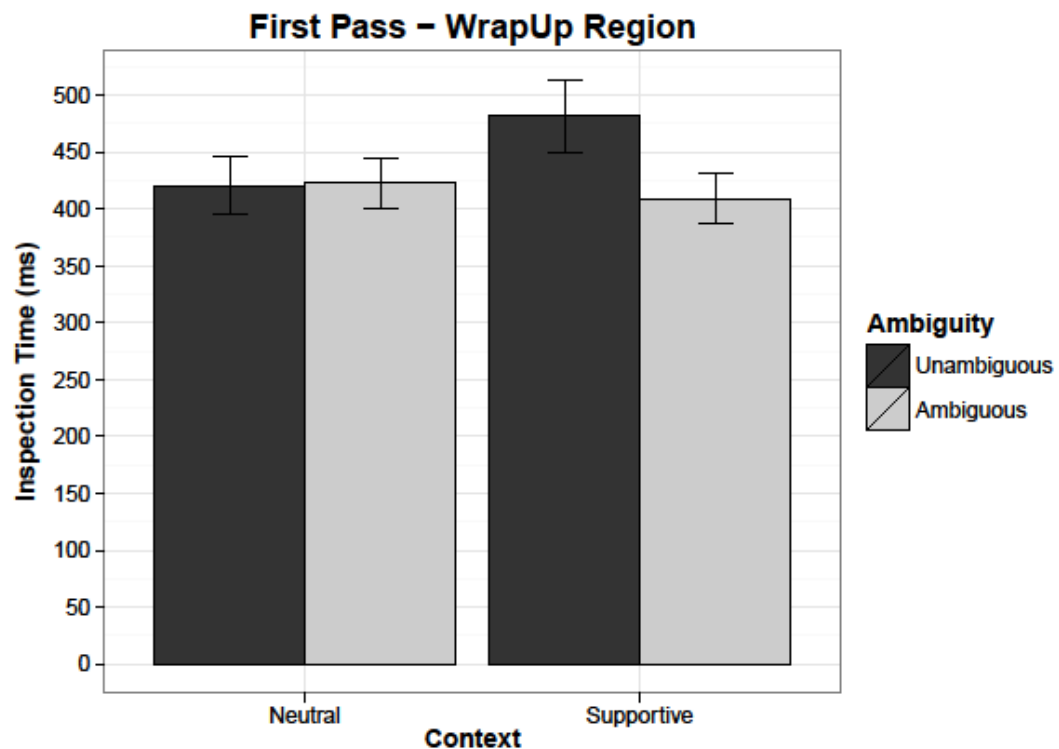


Figure 9.2: Experiment 4 - Mean inspection time for First Pass in the Critical Region. Error Bars display Standard Error. Context by Ambiguity Interaction.

This is difficult to explain as we would expect longer WrapUp times on the condition designed to be harder to process (Neutral / Ambiguous) and not in the 'easiest' condition (Supportive / Unambiguous). It is possible that the Supportive / Unambiguous condition is the only one in which normal wrapup occurs because all the other conditions are, when combined with cognitive load, difficult for some reason (whether it is because of ambiguity or due to a neutral context with limited discourse cohesion). If this is the case it is likely that the slow down for each other condition is being spread across the intervening regions so we are not able to observe it.

### Total Time in the Pre-Critical Region:

In Total Time, Pre-Critical Region we see a marginal interaction between Context and Ambiguity ( $p = 0.067^m$ ) (see Figure 9.3). It is worth being cautious about this observation as the region is very small (between 1 and 4 characters) and the interaction is only marginal. That said, we observe very little influence of Ambiguity when the Context is Supportive (Unambiguous = 299ms vs Ambiguous = 302ms,  $p > 0.1$ ), but a very significant slowdown for the Neutral / Ambiguous condition, as compared to the Neutral / Unambiguous condition (341ms vs 280ms,  $p = 0.005^{**}$ ). This implies that, when the Context is Supportive, Ambiguous words are processed equally as quickly as their Unambiguous controls, but when the Context is Neutral, this is not the case. The Ambiguous words take significantly longer to resolve than their Unambiguous controls.

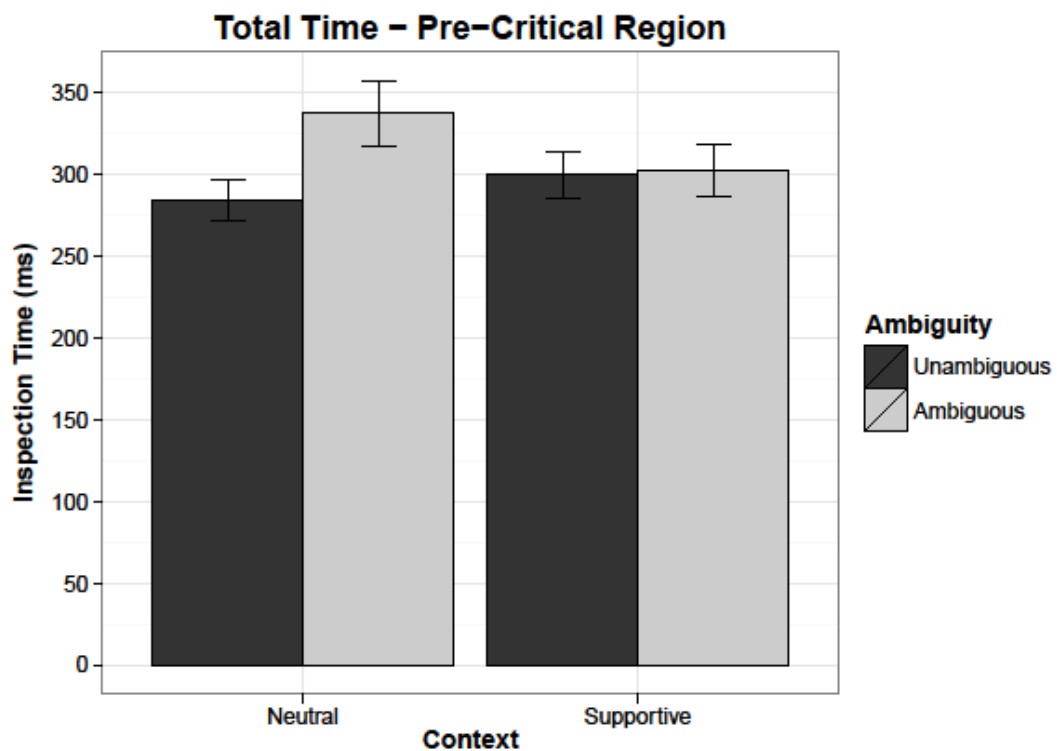


Figure 9.3: Experiment 4 - Mean inspection time for Total Time in the Pre-Critical Region. Error Bars display Standard Error. Context by Ambiguity Interaction - Marginal (0.076)

### Ambiguity by Load Interaction:

First pass of the Critical Region:

In the Critical Region we do not observe any interactions between Context and Ambiguity, but Ambiguity and Lexical Frequency do interact ( $p = 0.03 *$ ) (see Figure 9.4). When the Critical Word is Unambiguous we see longer fixations if it is a Low Frequency word as compared to a High Frequency word (258ms vs 224ms,  $p = 0.046 *$ ). Whereas there is very little difference in fixation durations when it is Ambiguous, regardless of Lexical Frequency (Unambiguous = 237ms vs Ambiguous = 247ms,  $p > 0.1$ ). As before, this is not expressly interesting in terms of the goals of this thesis, but it does indicate that low level factors are able to interact, even while the reader is under cognitive load.

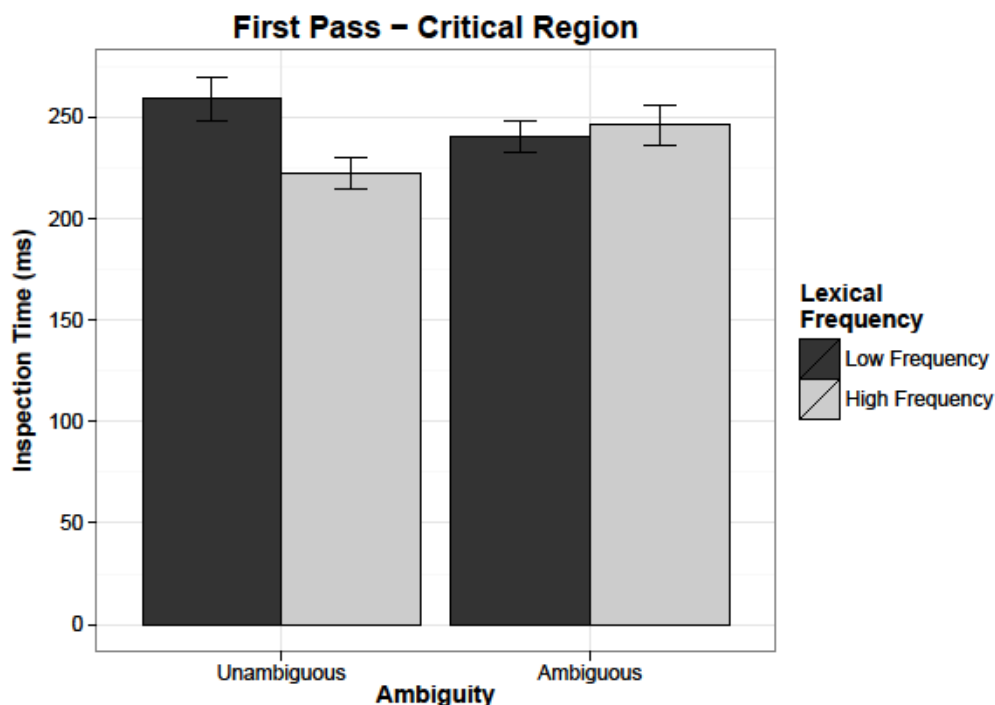


Figure 9.4: Experiment 4 - Mean inspection time for First Pass in the Critical Region. Error Bars display Standard Error. Ambiguity by Lexical Frequency Interaction.

### Summary of Main Effects:

The main effects are reported in the tables below, Table 9.2 reports Context effects, while Table 9.3 reports Ambiguity effects.

#### Context:

Context effects remain fairly similar to those observed in the NoLoad experiment. Early reading times appear mostly unaffected (with the exception of regression path times in the later regions), and in all cases a Neutral context causes longer re-reading times, and more regressions in to the critical regions (see Table 9.2). This is likely a hunt for propositional content which is lacking when the context is neutral. There is one exception to this pattern, which is the probability of fixating the combined Pre-Critical and Critical region. Neither region alone is significant (or even marginal) but combined we observe a greater overall tendency to skip the critical region if the context is neutral. This effect is likely spurious (an artefact of combining two regions) as it is hard to consider a theoretical reason for readers skipping the head noun when the context provides less information about it than when it is informative.

#### Ambiguity:

Table 9.3 shows a very slight alteration in the pattern of ambiguity main effects as compared to the NoLoad experiment. Here we see the presence of an ambiguity effect in the Wrapup region for first pass such that readers spent longer in wrapup if the critical word was unambiguous than if it was ambiguous. This is unexpected, but may be due to the ambiguity causing additional processing early on (partially supported by a marginal effect in the total time spent on the pre-critical region) so that upon reaching the end of the sentence, less wrapup is required. Additionally, we see a marginally higher probability of fixating the critical noun if it is ambiguous, which is supported by a lower probability of fixating the pre-critical region, if the upcoming word is ambiguous. Taken together, this indicates an attraction effect, in which the eyes are drawn to the critical noun more if it is ambiguous, causing the previous word to be skipped more often.

#### Frequency:

As before, Frequency appears to influence early reading times of the critical noun: The combined Pre-Critical and Critical region receives longer first pass reading times, and regression path times if the critical noun is of low frequency.

Table 9.2: Experiment 4 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if Context is Neutral or Supportive. Effects are also marked if Marginal.

<b>Context</b>	<b>Levels</b>	<b>Initial</b>		<b>Pre-Critical</b>		<b>Critical</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Pre-Crit + Crit</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Neutral	-0.19	681.2	0.38	228.1	-0.79	245.4	-1.30	346.9	0.97	424.6	-0.44	277.8
	Supportive		669.3		227.7		240.9		330.5		444.4		272.8
First Fixation	Neutral	-0.44	214.6	-0.76	224.8	-0.06	228.6	-1.18	246.6	1.09	247.0	-0.13	228.8
	Supportive		216.2		215.8		231.9		237.0		254.6		229.8
Regression Path	Neutral	1.01	715.6	-0.56	297.2	0.37	316.4	-2.00	925.6	-1.25	1740	-0.66	366.3
	Supportive		751.3		292.2		330.0	*	786.9		1626		362.9
Second Pass	Neutral	-2.62	466.1	-1.77	83.29	-2.39	147.3	-3.21	245.1	-1.34	134.3	-3.10	234.6
	Supportive	*	382.0	<sup>m</sup>	67.96	*	118.2	**	189.4		113.7	**	184.3
Total Time	Neutral	-1.99	1101	-0.57	306.5	-1.76	371.3	-3.03	571.5	-0.36	592.3	-2.49	482.6
	Supportive	*	1018		298.7	<sup>m</sup>	346.0	**	496.4		593.4	*	436.8
Probability of Fixation	Neutral	1.51	.93	0.69	.32	0.52	.59	0.70	.82	-1.98	.80	2.66	.77
	Supportive		.94		.34		.60		.83	*	.76	*	.82
Regressions Out	Neutral	1.00	.03	-1.54	.15	1.34	.12	-0.11	.34	-1.58	.82	-0.17	.13
	Supportive		.04		.11		.15		.34		.80		.14
Regressions In	Neutral	-1.45	.59	-1.65	.26	-2.30	.40	-2.17	.40	/	/	-2.57	.49
	Supportive		.55		.23	*	.34	*	.34		/	*	.43

Table 9.3: Experiment 3 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if the Critical Noun is **Ambiguous** or **Unambiguous**. Effects are also marked if **Marginal**.

<b>Ambiguity</b>	<b>Levels</b>	<b>Initial</b>		<b>Pre-Critical</b>		<b>Critical</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Pre-Crit + Crit</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unambiguous	-0.02	679.6	-0.26	228.6	-0.18	243.9	0.81	333.1	-1.97 *	454.1	-0.14	276.3
	Ambiguous		671.0		227.2		242.3		344.3		414.8		274.3
First Fixation	Unambiguous	-0.87	218.2	0.09	219.2	-0.09	229.6	0.59	239.4	-0.57	251.9	0.31	228.0
	Ambiguous		212.8		221.3		230.9		244.2		249.7		230.7
Regression Path	Unambiguous	0.44	731.1	0.91	278.1	-0.22	326.4	1.08	801.5	-0.56	1705	0.08	362.7
	Ambiguous		735.8		311.3		320.0		911.0		1660		366.5
Second Pass	Unambiguous	0.62	412.1	0.95	71.49	-0.36	135.9	0.02	218.4	0.58	118.1	-0.03	209.5
	Ambiguous		436.0		79.73		129.6		216.0		129.9		209.5
Total Time	Unambiguous	0.78	1048	1.84 m	290.0	-0.84	367.4	0.71	525.2	-0.64	604.3	0.50	454.6
	Ambiguous		1070		315.3		350.0		542.8		581.4		464.8
Probability of Fixation	Unambiguous	0.94	.93	-2.18 *	.35	1.90 m	.56	0.88	.82	-0.83	.79	0.31	.79
	Ambiguous		.94		.30		.62		.84		.77		.80
Regressions Out	Unambiguous	0.99	.03	0.61	.12	-0.07	.14	-0.09	.34	0.62	.80	-0.12	.14
	Ambiguous		.04		.14		.13		.33		.82		.13
Regressions In	Unambiguous	-0.20	.58	0.22	.24	-0.16	.37	-0.89	.38	/	/	0.13	.46
	Ambiguous		.57		.25		.37		.36		/		.46

## 9.4 Discussion

The addition of a cognitive load of this type does not appear to have a large impact on the normal processing of a sentence. We do not see a particular reduction in the use of context compared to the NoLoad version of the experiment. In terms of interactions, we see Ambiguity and Lexical Frequency interacting very early on (First Pass on the Critical word), which shows us that readers under load are still susceptible to the influence of low level lexical characteristics of words. This is not particularly interesting from our perspective as it does not impact the debate of SR vs. SLIR. Neither theory would predict a reduction in the influence of lexical properties. It is interesting, however, that we observe this occurring in the Load experiment slightly faster than in the NoLoad experiment: In the NoLoad experiment this occurred on the first fixation of the Spillover region, rather than the First Pass of the Critical region. This is a very slight temporal difference and so should be considered with caution, but it implies that readers under Load are slightly more attuned to low level factors than those not under Load. One possible explanation for this is that readers not under load are able to integrate higher level information into the discourse more readily and so need not attend low level factors as closely. This is simply conjecture however, and our experiments provide little evidence to impact this particular interpretation. There are hints of Low and High level factors interacting later on in the eye tracking record. In First Pass we observed a Context by Ambiguity interaction in the Wrapup region, which, while not patterning quite how the Capacity Constrained model would predict in either the SR or SLIR frameworks, shows that even while under Load, participants are still able to access higher level factors like context and bring them to bear on the processing of lexical ambiguity. In support of this we see a much more predicted pattern in the Total Time spent on the pre-critical region (the determiner in *'the* appendix') in which a Supportive context causes ambiguous words to be processed equally as quickly as their unambiguous controls, but a neutral context does not: The ambiguous words take significantly longer to resolve than their unambiguous controls. This interaction is only marginal however, so as mentioned above, it is worth being cautious of drawing any strong conclusions.

The principal finding of the NoLoad experiment was a Context by Ambiguity interaction in the probability of fixating the critical region. It appeared that readers were more likely to

fixate the critical region if it was ambiguous, but that this bias was much reduced when a supportive context was present. Under Load we see a very different pattern of results for this measure in this region. No Context by Ambiguity interaction is found, but we do observe a three way interaction between Context and Ambiguity and Frequency. No interaction is found between Context and Ambiguity for Low Frequency words, but an interaction is present for High Frequency words. The pattern for the high frequency words is not the same as that observed in the NoLoad experiment so it is not simply the case that Loading prohibits the use of contextual information to guide fixations when the upcoming critical word is of Low frequency.

The pattern of the high frequency data is such that ambiguous words are no more likely to be fixated than unambiguous words if the context is neutral (similar to the Low Freq data) but when the context is supportive there is a significant increase in the likelihood of fixation for an ambiguity, compared to an unambiguous control. This is almost the reverse of the pattern seen across both levels of frequency in the previous NoLoad experiment. One interpretation is that when a context is providing support for a high frequency word, the overall difficulty of para-foveal processing on that word is reduced enough that preview can provide identification and indicate to the reader the presence of an upcoming ambiguity. However, this is not the case for harder, low frequency words, and in either case, being under Load appears to inhibit the use of context to actually assist with the pre-processing of a word such that it doesn't require fixation (as in seen in the NoLoad experiment).

The interactions discussed above suggest that even under Load participants are still able to access and make limited use of contextual information when encountering lexical ambiguities. The data for early processing (FP, Pfix) however, seems difficult to interpret. The SR theory predicts a reduction or elimination of contextual influence on all processing such that the context effects observed in the NoLoad experiment would not be observed in the Load experiment. However, these main effects appear robust regardless of Load. In this way the data better supports the SLIR theory as it predicts no influence of Load on normal, interpretive processing. Both theories, however, predict a reduction or elimination of contextual effects on the processing of the lexical ambiguities themselves, which is not entirely supported by the data here. We see evidence that context is still being utilised to process ambiguity in several places, but in all cases but one (the marginal Total Time



Context by Ambiguity interaction on the pre-critical region), the nature of the interaction is not clear cut. Given that the nature of the differences between the experimental findings are somewhat unclear, it seems wise to check the robustness of these differences statistically before speculating on the reasons that they may differ.

In the next chapter we will present a statistical meta-analysis of the two experiments described above.

## Chapter 10: Combined Analysis of NoLoad and Load Experiments

### 10.1 Introduction

In the previous section we began to look at the differences between experiments and found that there were few differences in normal processing, but some differences in the nature of the interaction between context and ambiguity. In this chapter we will discuss what is observed statistically when a meta-analysis of two previous experiments is performed. First though, we will briefly isolate and describe the differences between the two experiments.

The tables below show the Context (Table 10.1), and Ambiguity (Table 10.2) main effects that differ between the NoLoad and Load experiments.

Table 10.1: Context Effects which were present / absent between experiments. NL = present in NoLoad experiment, L = present in Load experiment. x = not present.

<i><b>It was no good, <u>the appendix</u> would have to be removed.</b></i>						
	Initial	Pre-Crit	Critical	Spillover	Wrapup	Pre-Crit + Crit
First Pass						
First Fixation						
Regression Path					NL x	
Second Pass	NL x		x L			x L
Total Time	x L	NL x			NL x	x L
Probability of Fixation					x L	
Regressions Out						
Regressions In		NL x				

Table 10.2: Ambiguity effects which were present / absent between experiments. NL = present in NoLoad experiment, L = present in Load experiment. x = not present.

<i><b>It was no good, <u>the appendix</u> would have to be removed.</b></i>						
	Initial	Pre-Crit	Critical	Spillover	Wrapup	Pre-Crit + Crit
First Pass					x L	
First Fixation						
Regression Path						
Second Pass						
Total Time						
Probability of Fixation				NL x		
Regressions Out						
Regressions In						

Overall there is not much evidence indicating a reduction in contextual influence on normal processing. In Second Pass, Total Time and Regression Path for example, although there appear to be several differences in particular regions, the variance appears to balance out when considering the sentence as a whole. It is also the case that the vast majority of effects trend in the same direction.

Equally, while there was not much evidence of ambiguity in either experiment, what we did observe indicate similarities rather than differences between the two experiments. With the exception of the First Pass effect in the Wrapup region of the Load experiment we only see main effects of ambiguity in the Probability of Fixation measure, and all except one (the attraction effect in the pre-critical region of the Load experiment) pattern the same way (higher probability of fixation if ambiguous rather than unambiguous).

The main effects do not indicate much difference between Load groups. The interactions we observed are much less consistent however. The principal interaction we were seeking to replicate was the Context by Ambiguity interaction in the probability of fixating the critical word, seen in the NoLoad experiment. This interaction patterned such that readers were more likely to fixate the critical word if it was ambiguous compared to unambiguous, but that this bias was removed by the presence of a supportive context. In the Load experiment, we did not observe this interaction, but rather than there being no interaction

at all, there appeared to be a three way interaction between Context, Ambiguity and Lexical Frequency, in which Low Frequency words saw no influence of either Context or Ambiguity, but High Frequency words did. The interaction within the high frequency words however, patterned almost completely the opposite to the NoLoad interaction. This indicated that Load was having a reduced effect, rather than no effect at all.

As with the syntax chapter, at this stage, we have not conducted a statistical analysis to see whether the differences we note are reliable. Of particular interest will be to see what is going on with the Probability of Fixation interaction.

The data presented in the upcoming sections represents a combination of the data of each experiment, with experiment (Load) being added as a factor into the statistical analysis. If we observe interactions with Load we can conclude that the differences that we observe between the two experiments are real, and not a result of noise in the data. Whether the Probability of Fixation interactions interact with Load will be our principal interest.

## 10.2 Methods

### *Data Analysis*

As in the combined analysis in the Syntax section, a two level predictor was created to distinguish between Loading states (and thus experiments). Experiment 3 was indexed with 'NoLoad', while experiment 4 was indexed with 'Load'. The two datasets were then merged together. Analysis then proceeded as before, with the addition of the Loading State predictor. In addition, analysis of a further fixation measure was conducted to test the efficacy of the Loading manipulation. The analysis examined the total number of fixations made across the critical regions before the most rightwards of them was exited right. For this combined analysis, the critical regions are: pre-critical, critical, and spillover.

### 10.3 Results

*/It was no good /the /appendix /would have/ to be removed./*  
*/        2        / 3 /        4        /        5        /        6        /*

**1 = Context, 2 = Initial, 3 = Pre-Critical, 4 = Critical, 5 = Spillover, 6 = Wrapup**

#### **Number of Fixations:**

We measured the raw number of fixations made in and around the Critical Region as a crude but effective independent measure of the efficacy of the Cognitive Loading Manipulation. We found that participants under load made significantly fewer fixations in and around the Critical Region (Number of fixations: NoLoad = 3172 vs Load = 2750, Means: NoLoad = 2.49 vs Load = 2.20,  $p = 0.043^*$ ), when compared to those who were not cognitively loaded. As before, we view this as evidence that the digit span load manipulation is having an influence on eye movements and reading behaviour, but do not try to reason theoretically about what it might mean beyond that.

#### **Overview of Critical Findings:**

In this meta-analysis we seek to statistically test the differences observed in the two experiments reported above. We find the presence in the NoLoad experiment of the Context by Ambiguity interaction in the likelihood of fixating the critical word, and its subsequent absence in the Load experiment, to be statistically significant giving us a significant three-way interaction between Context, Ambiguity and Load (see Figure 10.1).

Additionally we observe a pair of three-way interactions between Context, Ambiguity, and Load in first pass reading times. The first is on the combined Pre-Critical + Critical region, but is difficult to interpret and only marginal. The second is on the spillover region and is significant. The NoLoad component of the spillover interaction replicates (one region downstream) the pattern observed by Duffy et al. (1988) but the Loaded side does not fit with the predictions of Miyake et al. (1994). See Figure 10.4 and section 10.4 for discussion.

### Context, Ambiguity and Load Interactions:

Probability of fixating the Critical Noun & and the Pre-Critical Region:

We observe several three way interactions between Context, Ambiguity, and Cognitive Load in the Probability of Fixation measure. We see one in the Initial Region ( $p = 0.002^{**}$ ), one in the Pre-Critical Region ( $p = 0.025^{*}$ ), and one in the Critical Region ( $p = 0.004^{**}$ ).

As in the first experiment (NoLoad, Context, and Ambiguity, section 8.3) we will not discuss the interaction in the Initial Region because the extremely limited number of observations of the region being Skipped (157 compared to 2431 cases that it was fixated) makes drawing conclusions about the influence of individual factors quite dubious. As before, we will present the results of the three-way interaction observed in the Critical Region first, followed by the Pre-Critical Region.

In the three way interaction between Context, Ambiguity, and Cognitive Load in the Critical Region ( $p = 0.004^{**}$ ) we observe a general reduction in the likelihood of fixation while under Load, compared to when not under Load (see Figure 10.1 & Table 10.3). Specifically, the Neutral / Ambiguous and the Supportive / Unambiguous conditions have significantly lower proportion of fixations while under Load (NoLoad = 70% vs Load = 59%,  $p = 0.021^{*}$ ) (NoLoad = 66% vs Load = 55%,  $p = 0.006^{**}$ ). We also see a minor reduction in the likelihood of fixation in the Neutral / Unambiguous condition, but this is not significant (NoLoad = 59% vs Load = 57%,  $p > 0.1$ ). The Supportive / Ambiguous condition remains the same across Load.

Table 10.3: Combined Analysis of Experiments 3 & 4 - Conditional Proportion of Fixations in % for Probability of Fixation: Critical Region. Context, Ambiguity, & Cognitive Load Interaction.

Context	Ambiguity	NoLoad	Load
Neutral	Unambiguous	59	57
Neutral	Ambiguous	70	59
Supportive	Unambiguous	66	55
Supportive	Ambiguous	64	64

When Loaded, the significant interaction between Context and Ambiguity that we observed in the NoLoad experiment ( $p = 0.009^{**}$ ) is no longer observed. Instead we observe a main

effect of Ambiguity ( $p = 0.036 *$ ) which is driven mostly by a significant pairwise effect across Ambiguity when the Context is Supportive (Unambiguous = 55% vs Ambiguous = 64%,  $p = 0.015 *$ ). In principal this means that participants are able to apply Context to guide eye movements when not under Load, but while under load their eye movements are guided only by Ambiguity. However, as we will see in the discussion of the Context, Ambiguity, and Frequency interaction (in the same measure and region) which will be discussed later in this section (see Figure 10.9); frequency may have a part to play in this interaction.

To anticipate that discussion in brief: We can see that the Loaded pattern here (panel 2) matches that of the High Frequency words in Figure 10.9. This indicates that while under Load participants can make use of contextual information, but only to support the processing of high frequency words (not low frequency words), and even then only to help indicate which upcoming words will require additional focus (rather than to help processing as we see in the comparable NoLoad group). This is likely due to the comparative ease of identifying and accessing high frequency words increasing the information gained in para-foveal uptake.

To summarise: When not Loaded we observe an increase of the likelihood of fixating the critical region if the context is Neutral and the critical word is Ambiguous and a reduction of the likelihood of fixating the critical word when it is Ambiguous and the context is Supportive. Thus some kind of para-foveal identification of an ambiguity, and the use of context to reduce the potency of the difficulty caused by it, appears to be occurring. Under Load these effects disappear, and this disappearance is driven by a lack of contextual influence on Low Frequency words. It appears that, under Load, high frequency ambiguities are able to be identified in the para-fovea (thus drawing fixations) but contextual facilitation of pre-processing of the kind seen in the NoLoad experiment, is not possible.

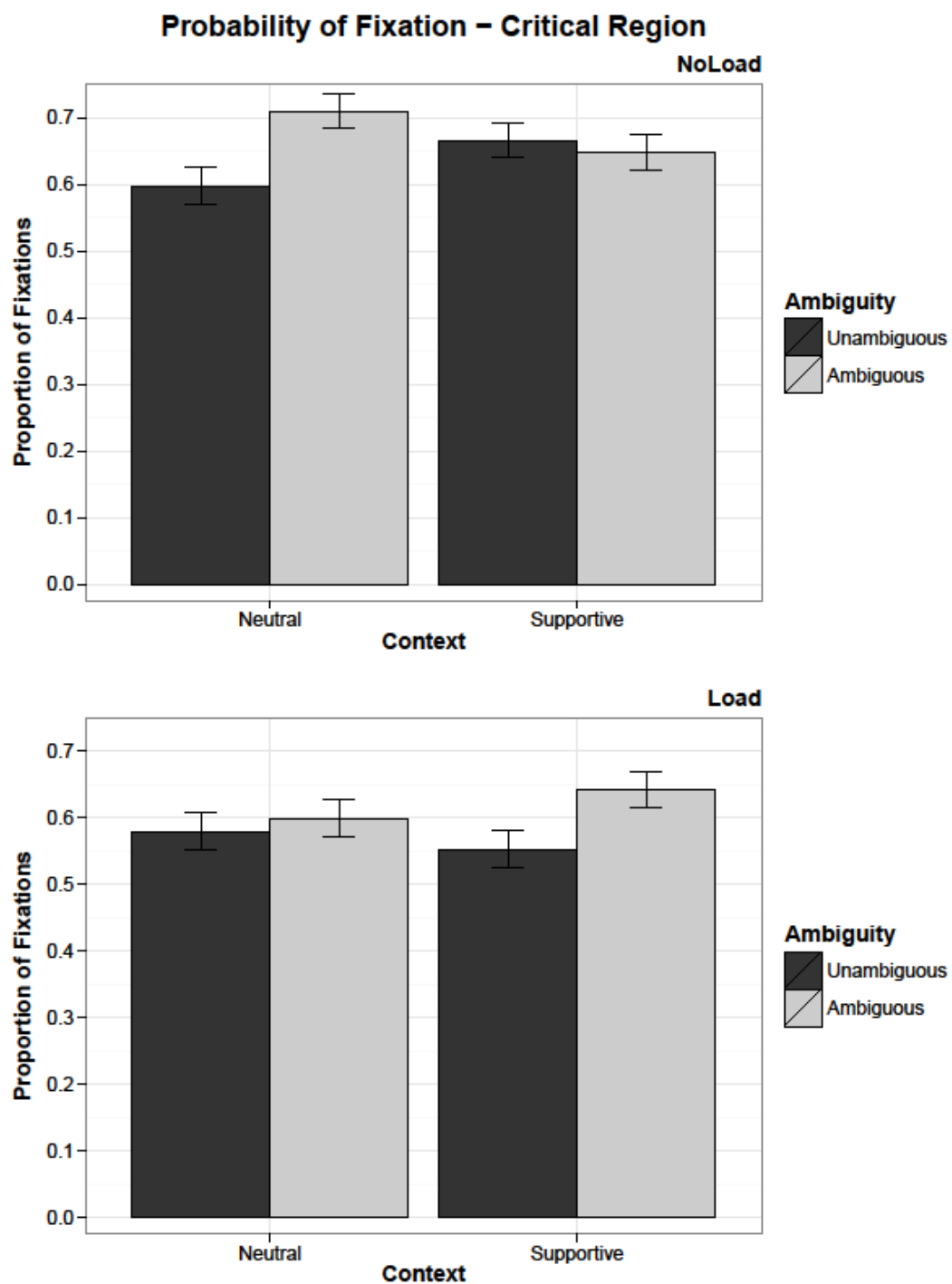


Figure 10.1: Combined Analysis of Experiments 3 & 4 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context, Ambiguity, and Cognitive Load Interaction.



As in the Critical Region, in the Pre-Critical Region we see a three way interaction between Context, Ambiguity, and Cognitive Load ( $p = 0.025^*$ ) which shows a general reduction in the likelihood of fixating the Pre-Critical Region while under Load (see Figure 10.2 & Table 10.4). Specifically the Neutral / Unambiguous condition which was one of the driving effects observed in the Results section of the first experiment (NoLoad, Context, & Ambiguity, section 8.3) has significantly reduced (NoLoad = 46% vs Load = 33%,  $p = 0.002^{**}$ ). In addition, both the Neutral / Ambiguous and Supportive / Ambiguous conditions have reduced marginally (NoLoad = 38% vs Load = 31%,  $p = 0.067^m$ ) (NoLoad = 39% vs Load = 30%,  $p = 0.061^m$ ). Conversely, the Supportive / Unambiguous condition has not reduced (in fact it has increased although not significantly (NoLoad = 36% vs Load 37 %,  $p > 0.1$ )).

Table 10.4: Combined Analysis of Experiments 3 & 4 - Conditional Proportion of Fixations in % for Probability of Fixation: Pre-Critical Region. Context, Ambiguity, & Cognitive Load Interaction.

Context	Ambiguity	NoLoad	Load
Neutral	Unambiguous	46	33
Neutral	Ambiguous	38	31
Supportive	Unambiguous	36	37
Supportive	Ambiguous	39	30

The significant pairwise differences that we observed in the NoLoad experiment (Neutral / Unambiguous > Supportive / Unambiguous,  $p = 0.005^{**}$ ) (Neutral / Unambiguous > Neutral / Ambiguous,  $p = 0.024^*$ ) are not observed when participants are under Load, but we do see a significant pairwise comparison across Ambiguity when the Context is Supportive (Unambiguous = 37% vs Ambiguous = 30%,  $p = 0.032^*$ ). It appears that we are seeing an attraction effect, as before. However, this time, rather than being triggered when the Context is Neutral and the upcoming Critical Word is Ambiguous (so, we see a higher fixation probability in this region for the Neutral / Unambiguous condition), it is being triggered when the Context is Supportive and Critical Word is Ambiguous (so it manifests in this region as a higher probability of fixation for the Supportive / Unambiguous condition).

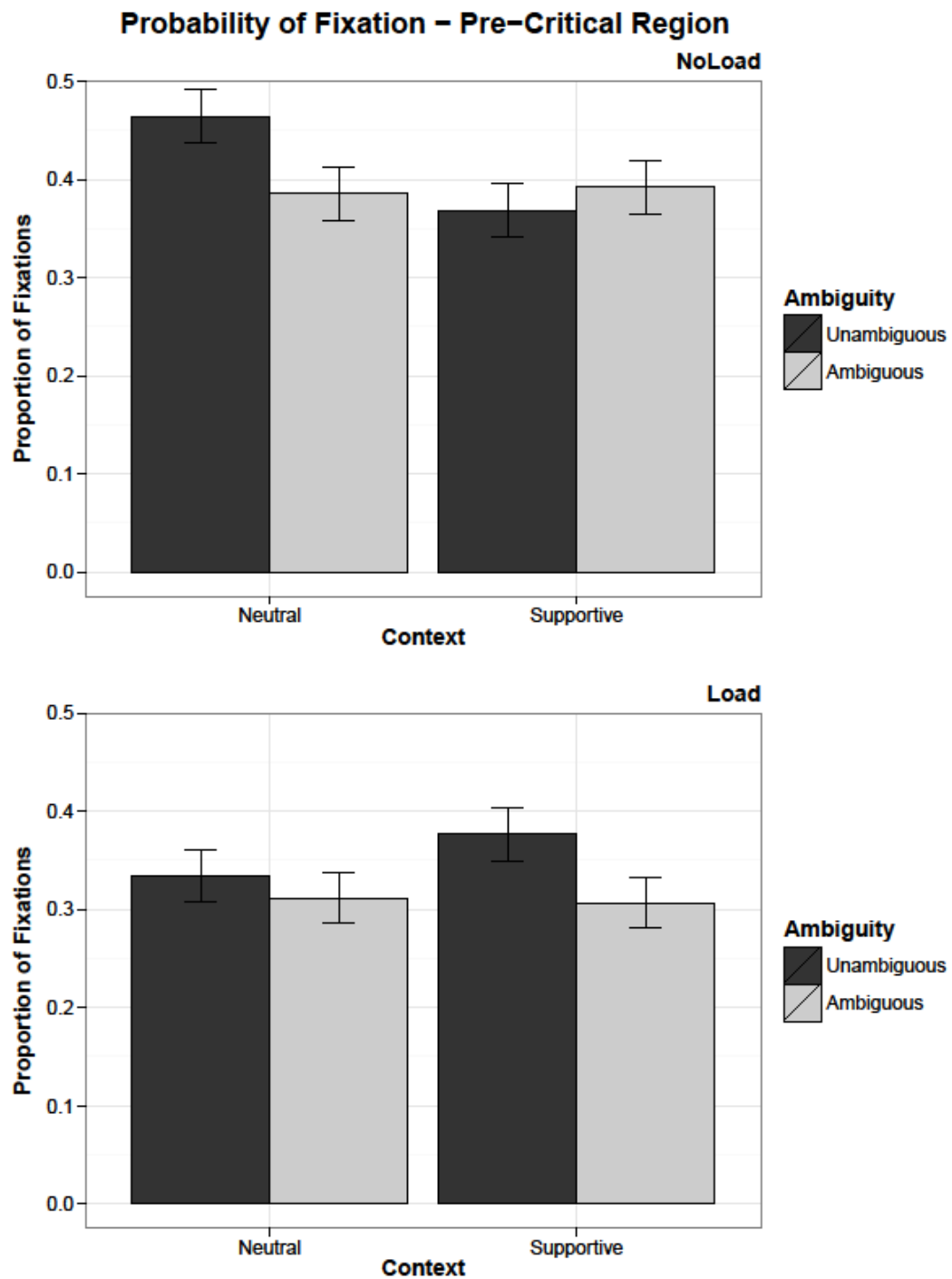


Figure 10.2: Combined Analysis of Experiments 3 & 4 - Proportion of Fixations for Probability of Fixation on the Pre-Critical Region. Error Bars display Standard Error. Context, Ambiguity, and Cognitive Load Interaction.

First Pass on the Pre-Critical + Critical region, and the Spillover region:

In First Pass we see Context by Ambiguity by Load interactions in the Combined Region: Pre-Critical & Critical, the Spillover Region, and the Wrapup Region, but only the one in the Spillover Region is significant, and the one in the Wrapup region is not particularly theoretically interesting and will not be discussed. (Pre-Critical & Critical:  $p = 0.073^m$ , Spillover:  $p = 0.037^*$ , WrapUp:  $p = 0.069^m$ ).

In the Combined Region: Pre-Critical & Critical, we see longer FP reading times for both the Neutral / Unambiguous and Supportive / Ambiguous conditions when not lot loaded as compared to when under load (310ms vs 272ms,  $p = 0.019^*$  and 289ms vs 266ms,  $p = 0.031^*$ ) (see Figure 10.3 & Table 10.5). No other pairwise comparisons are significant, and while the trends indicate numeric differences in the same direction for the remaining two conditions (Neutral / Ambiguous & Supportive / Unambiguous), the differences are much less dramatic (see Table 10.5). It appears then, that being under Cognitive Load decreases First Pass reading times across the board but influences certain conditions more strongly. It is worth being cautious about reading too much into this interaction, however, as this is both a combination of two regions (which may be muddying the waters) and only Marginal.

Table 10.5: Combined Analysis of Experiments 3 & 4 - Conditional Mean fixation durations in ms for First Pass: Combined Region: Pre-Critical & Critical. Marginal Context, Ambiguity, & Load Interaction.

Context	Ambiguity	NoLoad	Load
Neutral	Unambiguous	310	272
Neutral	Ambiguous	290	283
Supportive	Unambiguous	288	277
Supportive	Ambiguous	289	266

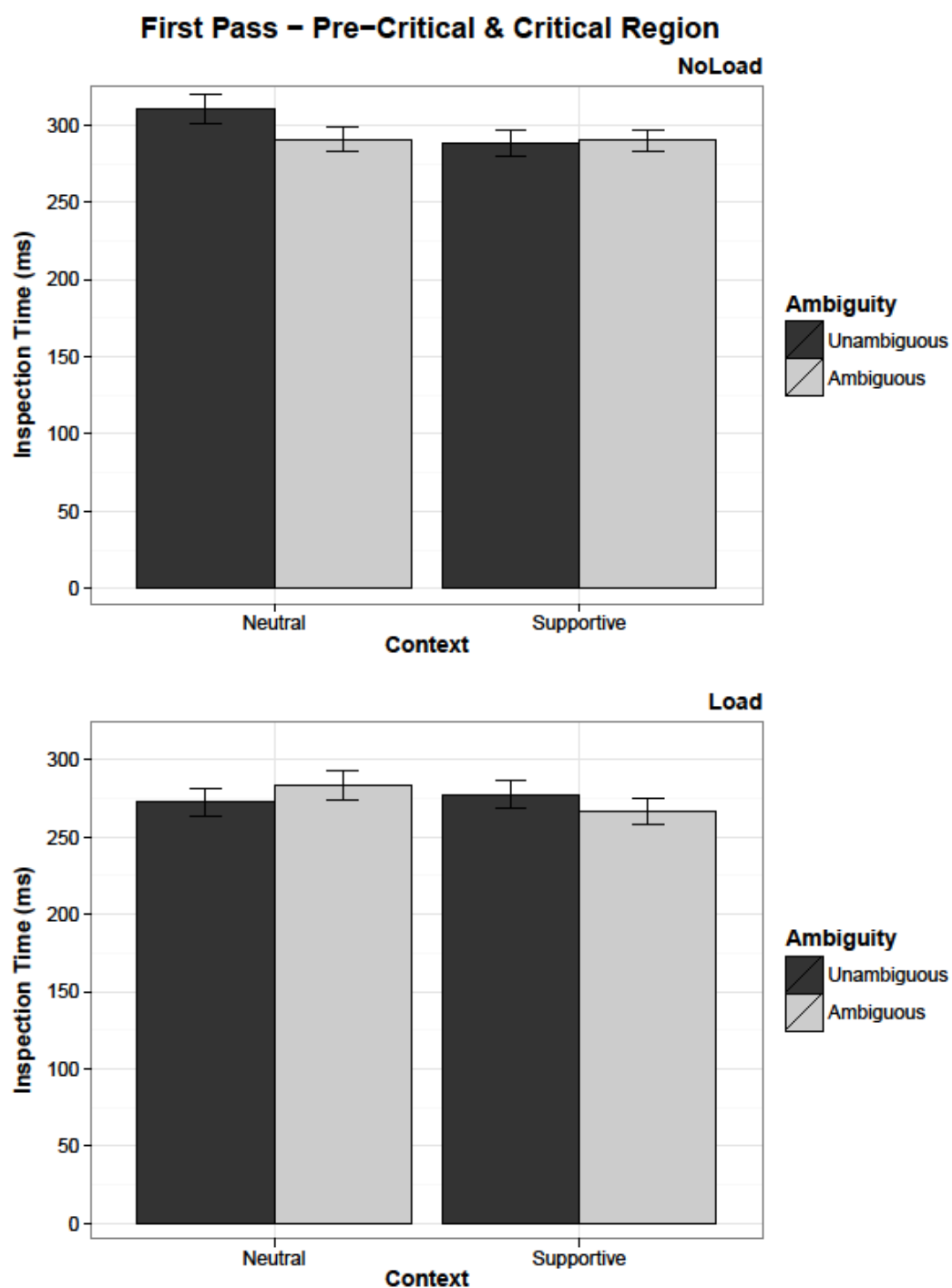


Figure 10.3: Combined Analysis of Experiments 3 & 4 - Mean inspection time for First Pass on the Combined Region: Pre-Critical & Critical. Error Bars display Standard Error. Marginal Context, Ambiguity, & Load Interaction.

In the significant interaction in the Spillover Region we observe a Marginal difference across Ambiguity when the Context is Neutral and there is no cognitive loading: We see longer FP fixation durations when the critical word is Ambiguous, compared to when it is not (372ms vs 336ms  $p = 0.06^m$ ). This effect disappears when under Cognitive Load (346ms vs 246ms,  $p > 0.1$ ) (see Figure 10.4 & Table 10.6). When the Context is Supportive there are no significant pairwise effects of Ambiguity for NoLoad (354ms vs 357ms,  $p > 0.1$ ) or Load (320ms vs 338ms,  $p > 0.1$ ). Numerically however, when under Load the Ambiguous condition appears to elicit longer reading times, compared to the Unambiguous condition but, when not loaded, the numbers are essentially no different. We observe a significant pairwise difference in the Supportive / Unambiguous condition across Load: When participants are not under load we see longer FP reading times as compared to when they are Loaded (354ms vs 320ms,  $p = 0.021^*$ ). To summarise: When not under Cognitive Load, if the Context is Neutral we see longer reading times for Ambiguous words (marginal), while if the Context is Supportive, Ambiguity appears to make no difference. When under Load however, if the Context is Neutral, Ambiguity appears to make no difference, while if the Context is Supportive we see shorter reading times for Unambiguous words (numerically). The implication is that readers not under Load are exhibiting expected behaviour (difficulty in processing an ambiguity that is not supported by a context), while those under Load appear to find all conditions equally difficulty except the 'easiest' one (Unambiguous, Supported).

Table 10.6: Combined Analysis of Experiments 3 & 4 - Conditional Mean fixation durations in ms for First Pass: Spillover Region. Context, Ambiguity, & Load Interaction.

Context	Ambiguity	NoLoad	Load
Neutral	Unambiguous	336	346
Neutral	Ambiguous	372	346
Supportive	Unambiguous	354	320
Supportive	Ambiguous	357	338

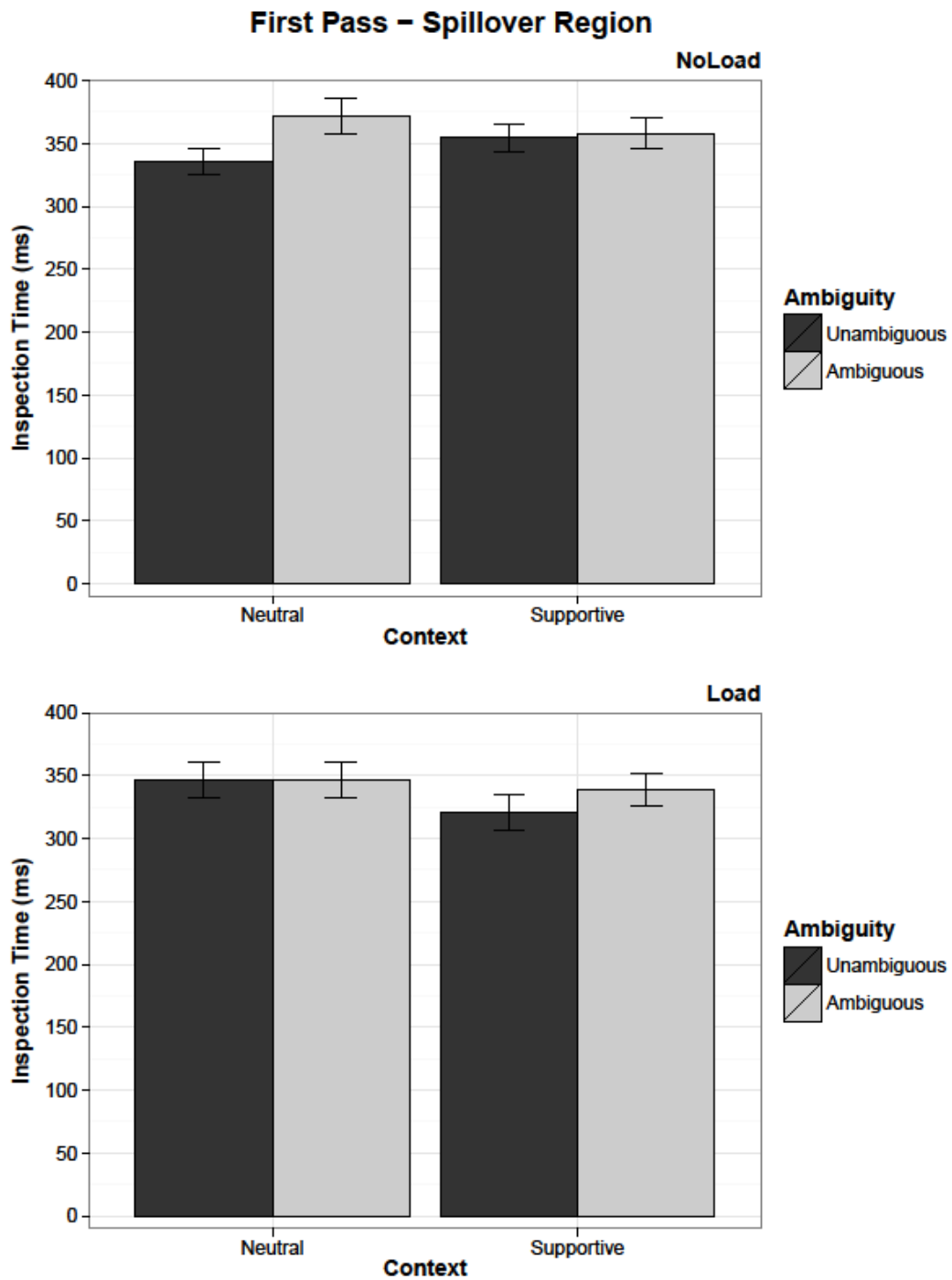


Figure 10.4: Combined Analysis of Experiments 3 & 4 - Mean inspection time for First Pass on the Spillover Region. Error Bars display Standard Error. Context, Ambiguity, & Load Interaction.

### Total Time spent on the Pre-Critical Region:

In Total Time, there is a three way Context, Ambiguity, & Load interaction ( $p = 0.019 *$ ) (see Figure 10.5 and Table 10.7). We see that when participants are not under Load, and when the Context is Neutral, there is a statistically significant increase in reading times if the upcoming critical word is Unambiguous rather than Ambiguous (325ms vs 291ms,  $p = 0.039 *$ ). If the context is Supportive we see a slight numerical difference in the opposite direction (longer reading times if Ambiguous rather than Unambiguous) but this is not significant (Unambiguous = 283ms vs Ambiguous 288ms,  $p > 0.1$ ). When Loaded, the Supportive condition shows the same pattern as when not Loaded (Unambiguous 299ms vs 302ms,  $p > 0.1$ ) but when the Context is Neutral, the significant pattern we saw when not Loaded is reversed (Unambiguous = 280ms vs Ambiguous = 341ms,  $p = 0.005 **$ ). In addition, we observe a significant pairwise effect across Load for the Neutral / Unambiguous condition (NoLoad = 325ms vs Load: 280ms,  $p = 0.004 **$ ). There are no other significant or marginal pairwise results.

Table 10.7: Combined Analysis of Experiments 3 & 4 - Conditional Mean fixation durations in ms for Total Time: Pre-Critical Region. Context, Ambiguity, & Load Interaction.

Context	Ambiguity	NoLoad	Load
Neutral	Unambiguous	325	280
Neutral	Ambiguous	291	341
Supportive	Unambiguous	283	299
Supportive	Ambiguous	288	302

Comparing the means of all conditions observed while not under Load it becomes clear that the Neutral / Unambiguous condition is causing participants to slow down (rather than any other conditions speeding up). It is unclear as to why this should be the case. Doing the same for Loaded participants however, we see that the Neutral / Ambiguous condition is causing participants to slow down (rather than any other conditions speeding up). We know from the Number of Fixations analysis (reported at the start of this section) that participants make fewer fixations while under Load, so it is possible that this display of difficulty is reflected in the Pre-Critical region (rather than the Critical region) because participants land on the Pre-Critical region and, rather than making a small saccade

forwards, are more inclined to engage in partial or total para-foveal processing in order to pre-process or process the Critical word. This is not observable in First Pass, however, so we must consider that it is only part of what is generating this effect. At least part of this observed effect will be generated by Second Pass reading times, in which participants are more likely to engage in re-reading if the initial reading was difficult in some way.

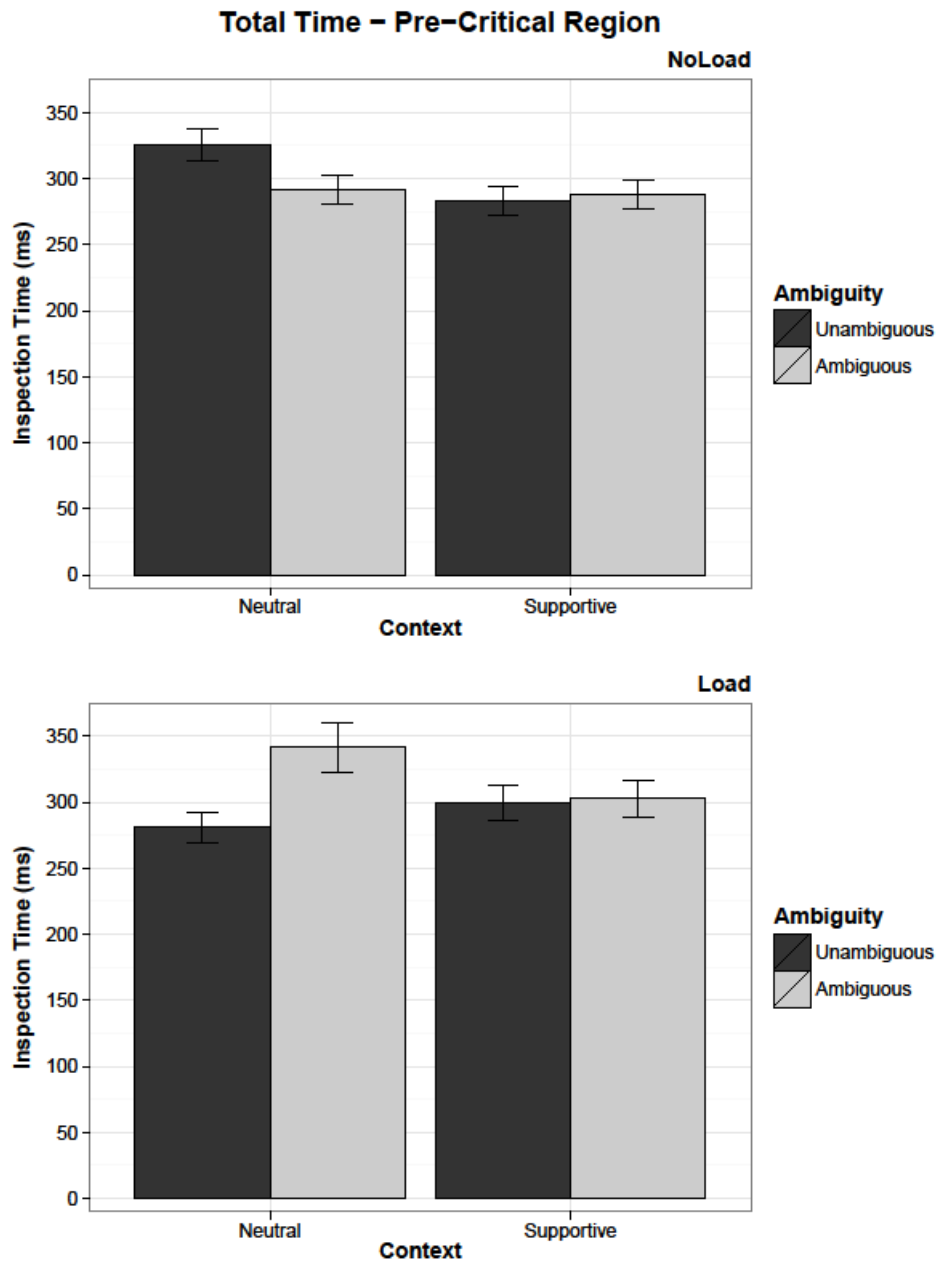


Figure 10.5: Combined Analysis of Experiments 3 & 4 - Mean inspection time for Total Time on the Pre-Critical Region. Error Bars display Standard Error. Context, Ambiguity, & Load Interaction.



### **Other Interactions:**

In the sections below we will discuss any relevant interactions involving Context, Ambiguity, or Cognitive Load. Space constraints prevent us from detailing all interactions exhaustively, so several irrelevant interactions will not be discussed, but these (mostly involving the control factor frequency) are listed in Table 10.8 below. Those involving three way interactions between Context, Ambiguity and Load will be very briefly summarised below Table 10.8.

To anticipate the main points of our results we find:

Context by Load interactions:

Probability of Fixating the Pre-Critical + Critical, and Wrapup Regions:

It appears that the probability of fixating the Combined region is not mitigated by context for those not under Load. For those under Load, however, supportive contexts appear to increase the likelihood of fixating the Pre-Critical + Critical region when compared to Neutral contexts (see Figure 10.6).

For the Wrapup region we see the same non-mitigation for NoLoad readers, but Loaded readers are more likely to fixate the final Wrapup region if the context was neutral, compared to supportive (see Figure 10.7).

Load by Lexical Frequency interactions:

Probability of Fixating the Critical Region:

We discuss this frequency interaction as it reflects on the differences observed between the NoLoad and Load experiments. In the NoLoad experiment we observed no Context by Ambiguity interaction with Frequency, but in the Load experiment the three factors did interact. For participants under Load, this Load by Frequency interaction shows the reverse pattern for participants to that which we saw in the NoLoad experiment: Under Load, High Frequency words do not differ in their processing to Low Frequency words (see Figure 10.8).

### Context, Ambiguity and Frequency Interactions:

#### Probability of Fixating the Critical Region:

We see an almost exact replication of this interaction in the combined analysis as we did in the Load experiment. Low frequency words are not influenced by context or ambiguity, but High frequency words are. The interaction seen for High frequency words does not follow the same pattern as the context by ambiguity interactions we have seen previously however. We refer the reader to section 9.4 and Figure 9.1 for discussion, but also summarise our interpretation (see Figure 10.9).

Table 10.8: Combined Analysis - Interactions involving Frequency which are not reported in the Results section.

Factors	Measure	Region
Freq + Ambiguity	Regressions In	Critical
Freq + Load	Second Pass	Spillover
	Probability of Fixation	Critical
Freq + Ambiguity + Load	Second Pass	Wrapup
Freq + Context + Load	Probability of Fixation	Wrapup
Context + Ambiguity + Load	First Pass	Wrapup

### Context, Ambiguity and Load interaction (not discussed):

#### First Pass for the WrapUp Region:

In First Pass for the WrapUp region all conditions across Context, Ambiguity, and Load are equal except for the Supportive / Unambiguous condition, which is comparatively slowed. In the previous (spillover) region we saw this condition being speeded so it is likely that this is spillover processing from the spillover region.

### Context by Load Interactions:

Probability of Fixating the Pre-Critical + Critical Region:

We observe a Context and Cognitive Load interaction in the Combined Region: Pre-Critical & Critical ( $p = 0.009^{**}$ ).

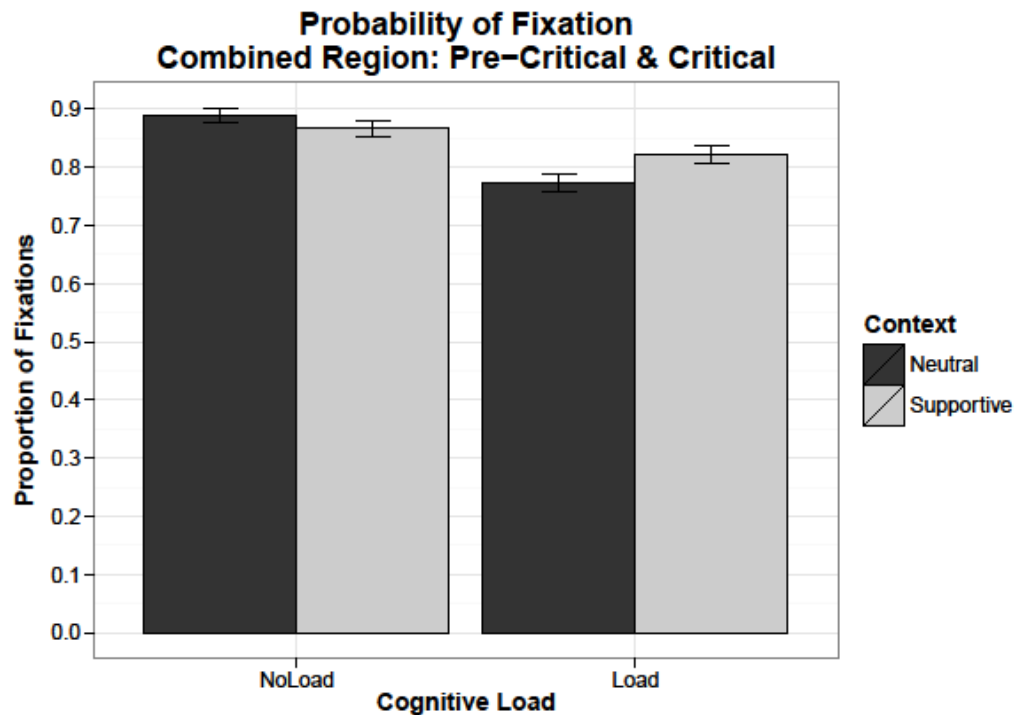


Figure 10.6: Combined Analysis of Experiments 3 & 4 - Proportion of Fixations for Probability of Fixation on the Combined Region: Pre-Critical & Critical. Error Bars display Standard Error. Context and Cognitive Load Interaction.

The Combined Region: Pre-Critical & Critical is a measure of the proportion of trials in which either the Pre-Critical or the Critical Regions were fixated on the first pass. When not under Load we observe a numerically higher likelihood of fixating the Combined Region if the context is Neutral, than if it is Supportive, but this is not significant (89% vs 86%,  $p > 0.1$ ) (see Figure 10.6). However, when under Load we see a significant pairwise difference in the opposite direction (Neutral = 77% vs Supportive 82%,  $p = 0.013^{*}$ ). We also observe a significant pairwise difference across Load when the Context is Neutral (NoLoad = 89% vs Load = 77%,  $p < 0.000^{***}$ ). When the Context is Supportive we see no such effect (NoLoad = 86% vs Load = 82%,  $p > 0.1$ ). So it appears that people are more likely to fixate the Combined Region when they are not Loaded than when they are Loaded but that the

context makes no difference to those not under Load. For those under Load, supportive contexts appear to increase the likelihood of fixating the Combined region. It is worth noting that for three reasons this interaction should be treated cautiously: Firstly, these effects are very close to ceiling and so might the proportion of data representing the instance in which the region was skipped is very small. Secondly, this is a combined region which contains a determiner and the head noun, and the measure is the probability of fixating either. There is, therefore, a possibility that we are seeing a mixture of the behaviours associated with fixating both of those regions individually. And thirdly, this interaction is superseded by a Context, Ambiguity, and Load interaction in the likelihood of fixating the critical region, which means that ambiguity will be having some effect, just not enough to create a statistically significant interaction in the combined region.

#### Probability of Fixating the WrapUp Region:

In the Wrapup Region ( $p = 0.046 *$ ), when Not Loaded we observe a small numerical difference across Context but this is not significant (Neutral = 81% vs Supportive = 82%,  $p > 0.1$ ) (see Figure 10.7). Conversely, when Loaded, we see a significant pairwise difference across Context (Neutral = 80% vs Supportive = 76%,  $p = 0.028 *$ ). We also see a significant pairwise difference across Load when the Context is Supportive (NoLoad = 82% vs Load = 76%,  $p = 0.025 *$ ), whereas when the Context is Neutral we do not (NoLoad = 81% vs Load = 80%,  $p > 0.1$ ). We see then, that when under Load, people are more likely to fixate the final Wrapup Region if the Context did not provide much propositional content. Additionally, when the Context is Supportive people are more likely to fixate the Wrapup region if they are not under Load than if they are. It is worth being cautious about this interaction, however, as the values are getting very close to ceiling, which isn't unexpected as it is the final region of the text.

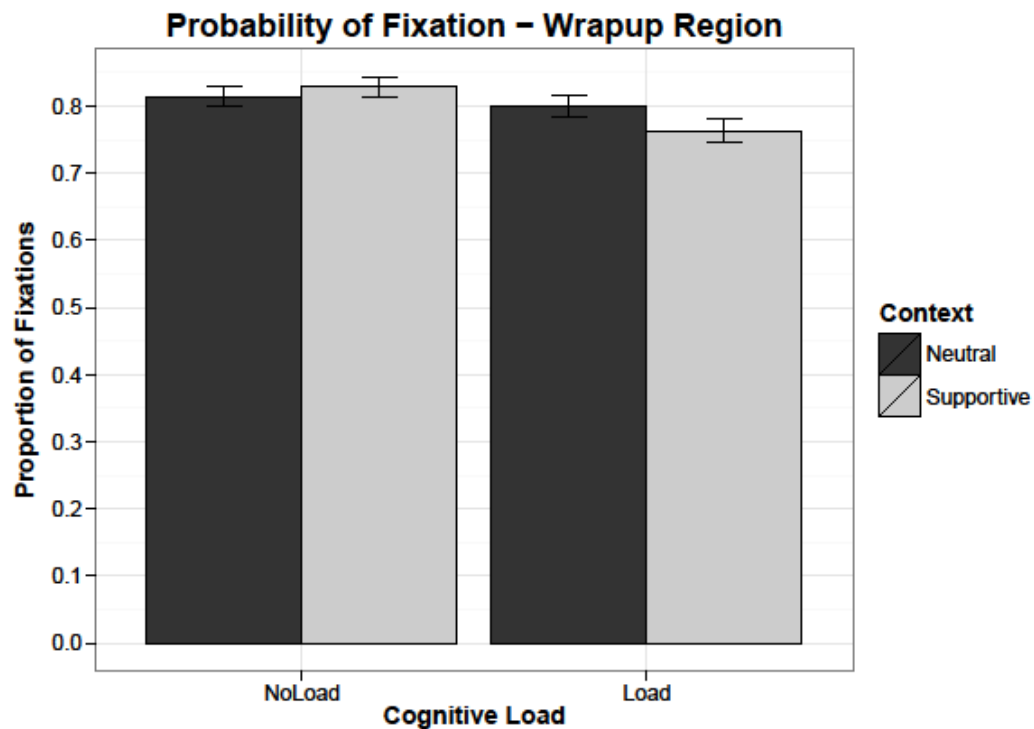


Figure 10.7: Combined Analysis of Experiments 3 & 4 - Proportion of Fixations for Probability of Fixation on the Wrapup Region. Error Bars display Standard Error. Context and Cognitive Load Interaction.

#### Load by Frequency Interaction:

Probability of Fixating the Critical Region:

We observe a Load by Frequency interaction ( $p = 0.012 *$ ) in the probability of fixating the Critical region (see Figure 10.8), which we report because of the differences seen in the processing of words of different Lexical Frequency in this measure and region between the previous NoLoad and Load experiments. We observed a Context by Ambiguity interaction in the NoLoad experiment (see Figure 8.1), which did not interact with frequency, but in the Load experiment we saw a differently patterning Context, Ambiguity, and Frequency interaction (see Figure 9.1). In this two way Load by Frequency interaction we see a numerically higher likelihood of fixating the critical region for those not under Load, if the critical word is of high frequency compared to low frequency, although this pairwise comparison is not significant (Low Freq= 63% vs High Freq = 69%,  $p > 0.1$ ). Under Load there

is no difference between Frequency conditions (Low Freq= 61% vs High Freq = 58%,  $p > 0.1$ ). There is a significant pairwise comparison across Load such that High Frequency words in the NoLoad condition are more likely to be fixated than High Frequency words in the Load condition (69% vs 58%,  $p = 0.004^{***}$ ) which implies that High frequency words are treated differently to Low frequency words when not under Load but, contrary to what we saw in the three way interaction of the Load experiment, no differently when under Load. Low frequency words do not differ across Load (NoLoad = 63% vs Load = 61%,  $p > 0.1$ ). Overall this interaction implies that the processing of Low frequency words does not differ across Load, but High frequency words are, in general, fixated more by readers not under Load than by those who are.

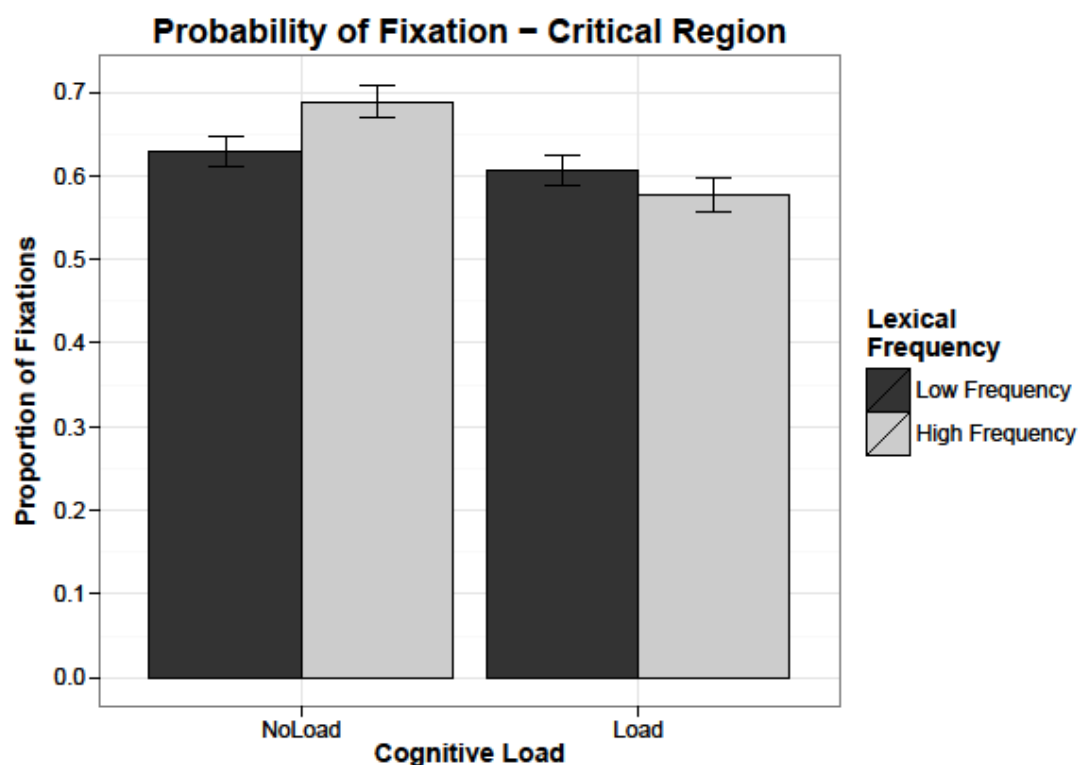


Figure 10.8: Combined Analysis of Experiments 3 & 4 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Lexical Frequency and Cognitive Load Interaction.

### **Context, Ambiguity, and Frequency Interaction:**

#### **Probability of Fixating the Critical Region:**

Here we see a Context, Ambiguity and Frequency interaction ( $p = 0.034 *$ , see Figure 10.9) which is a near exact replication of the one we saw in this measure and region for the Load experiment (see Figure 9.1 and section 9.4 for discussion).

However, as much as this interaction represents the Load experiment (with little influence of the NoLoad experiment), the pattern of the Context by Ambiguity interaction for the high frequency words (see panel 2) is not the same as that observed for all words in the Context by Ambiguity interaction in the NoLoad experiment (see Figure 8.1). Given that these patterns differ, it cannot simply be the case that Loading outright prohibits the use of contextual information to guide fixations when the upcoming critical word is of Low frequency, but does not affect its use in guiding fixations around high frequency words.

Rather, a potential explanation is that the contextually facilitated para-foveal processing which extends beyond simple identification (as suggested previously) is inhibited by Load; but when the context is providing support for a high frequency word, the overall difficulty of para-foveal processing on that word is reduced enough that preview can still provide identification and indicate the presence of an upcoming ambiguity. However, this is not the case for harder, low frequency words, which remain too difficult to be identified in the para-fovea. This will be discussed in greater depth in the next section (10.4).

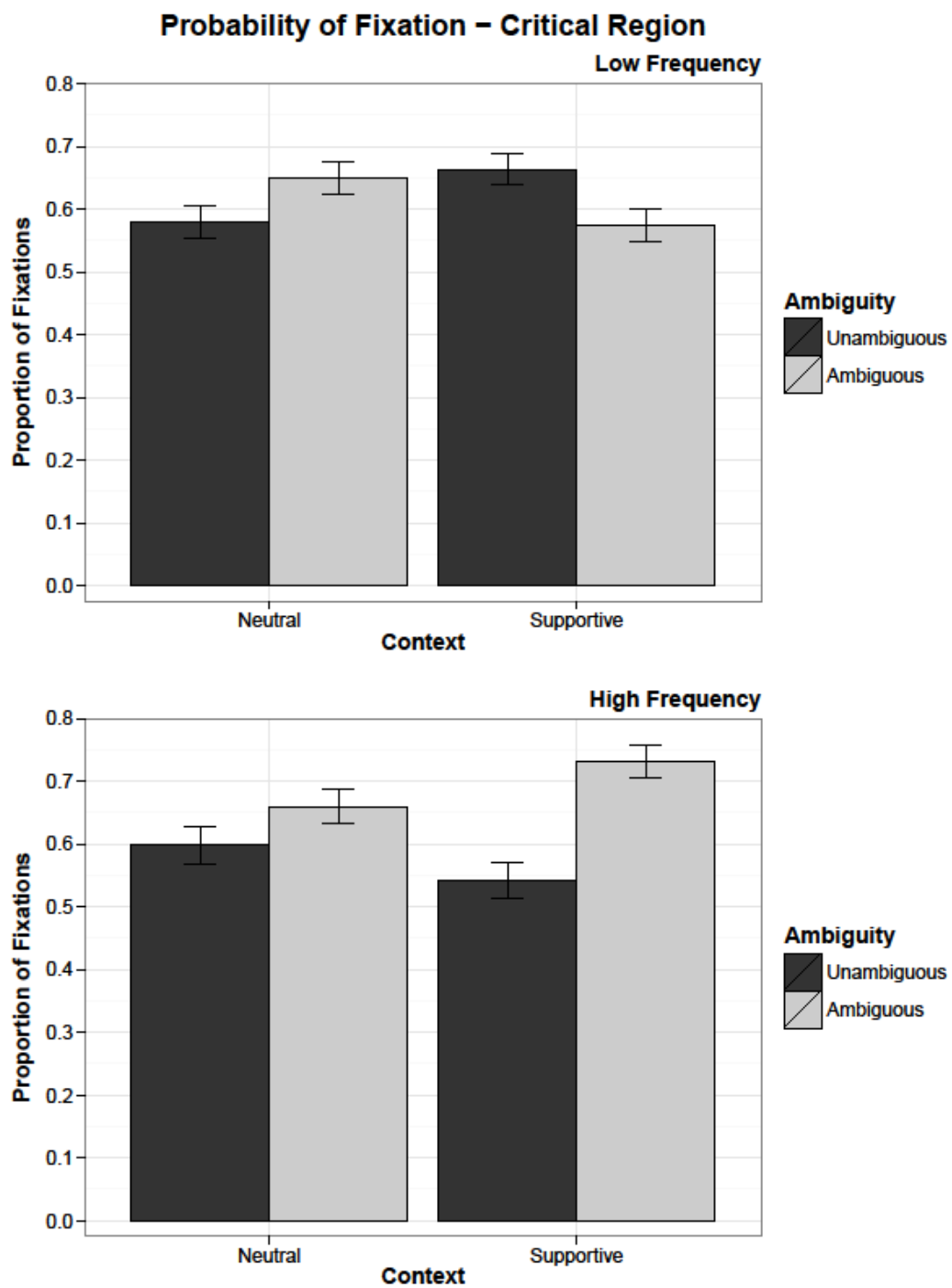


Figure 10.9: Combined Analysis of Experiments 3 & 4 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context, Ambiguity, and Frequency Interaction.



### **Combined Analysis, Summary of Main Effects:**

The main effects are reported in the tables below. Table 10.9 reports Context effects, while Table 10.10 reports Ambiguity effects, and Table 10.11 reports Load effects.

#### **Context:**

Table 10.9 displays the main effects of context for this combined analysis. As in the individual NoLoad and Load experiments, the overriding trend is for a supportive context to facilitate processing in the later reading measures and fixation measures (shorter re-reading times (SP, TT) and fewer regressions to previous regions (RI)). The one effect which went counter to this trend (Pfix, Pre-Critical + Critical, Load experiment) is no longer observed.

#### **Ambiguity:**

The main effects of ambiguity in this combined analysis are minimal (see Table 10.10) as they were in the individual experiments. The only two observed here are opposing probability of fixation effects for the pre-critical and critical regions. The effect on the Pre-Critical region (in which a higher likelihood of fixation is observed if upcoming noun is unambiguous) was present in the Load experiment, while the effect on the Critical region (which patterns the opposite) was present in the NoLoad experiment (although was marginal in the Load experiment as well). Together, they appear to suggest an attraction effect, similar to as reported in the Load experiment, such that the eyes are drawn to the critical noun more often if it is ambiguous, and often this involves skipping the preceding word.

#### **Load:**

The overall pattern of the Load main effects (seen in Table 10.11) suggests that participants under Load are reading more quickly (FP; Initial, Pre-Critical +Critical), engaging in less mid-sentence regressions (RO; Critical), skipping regions more (Pfix; Pre-Critical, Critical, Spillover), and engaging in more regressive and longer re-reading behaviours after reaching the end of the sentence (RO; WrapUp, RP; Wrapup, SP; Initial) than those not under Load. This implies speeded initial readings which are shallower and less thorough than their not Loaded counterpart's, as well as a need to re-read and reassess for proper comprehension.

Table 10.9: Combined Analysis of Experiments 3 & 4 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if Context is **Neutral** or **Supportive**. Effects are also marked if **Marginal**.

<b>Context</b>	<b>Levels</b>	<b>Initial</b>		<b>Pre-Critical</b>		<b>Critical</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Pre-Crit + Crit</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Neutral	0.82	770.8	0.19	232.8	-1.04	236.8	-0.59	353.6	-0.37	377.6	-1.13	288.0
	Supportive		675.3		227.9		242.2		338.7		434.5		279.6
First Fixation	Neutral	-1.22	218.0	-0.16	222.4	-0.54	228.1	-0.42	244.4	-0.44	243.6	-0.22	228.9
	Supportive		216.6		220.8		227.3		240.9		246.5		228.8
Regression Path	Neutral	1.57	757.7	-0.83	304.0	-0.08	232.3	-2.52	819.6	-2.47	1544	-1.06	378.7
	Supportive		788.0		299.3		328.3	*	724.9	*	1434		372.1
Second Pass	Neutral	-2.41	383.1	-3.15	86.17	-2.47	131.9	-4.30	239.1	-2.01	138.0	-3.18	214.7
	Supportive	**	327.5	**	68.20	*	112.4	***	190.7	*	115.9	**	176.3
Total Time	Neutral	1.91	1057	-1.97	306.0	-1.82	354.8	-3.92	570.5	-2.14	579.3	-4.07	479.8
	Supportive	<sup>m</sup>	1015	*	291.6	<sup>m</sup>	339.8	***	511.9	*	551.9	***	435.9
Probability of Fixation	Neutral	0.42	.94	-0.79	.37	0.47	.62	-0.98	.86	-0.87	.81	0.68	.83
	Supportive		.94		.35		.63		.86		.79		.84
Regressions Out	Neutral	0.89	.03	-1.43	.17	0.98	.16	-0.84	.34	-1.85	.79	0.05	.15
	Supportive		.03		.14		.18		.32	<sup>m</sup>	.76		.15
Regressions In	Neutral	-2.32	.54	-2.76	.29	-3.48	.39	-2.08	.39	/	/	-3.97	.49
	Supportive	*	.50	*	.24	**	.33	*	.35	/	/	***	.42

Table 10.10: Combined Analysis of Experiments 3 & 4 - Main Effects of Ambiguity. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if the Critical Noun is Ambiguous or Unambiguous. Effects are also marked if Marginal.

<b>Ambiguity</b>	<b>Levels</b>	<b>Initial</b>		<b>Pre-Critical</b>		<b>Critical</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Pre-Crit + Crit</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unambiguous	-0.30	723.1	-0.77	233.9	-0.45	239.6	1.40	338.1	-1.63	416.5	-0.46	286.2
	Ambiguous		723.0		226.8		239.4		354.2		395.5		281.4
First Fixation	Unambiguous	-0.88	219.3	-0.49	223.4	-0.54	227.5	0.91	240.0	-0.44	244.9	-0.18	228.4
	Ambiguous		215.3		219.8		227.9		245.3		245.2		229.3
Regression Path	Unambiguous	-0.27	774.3	0.38	293.7	-0.40	329.1	0.60	750.5	-0.71	1513	-0.11	376.1
	Ambiguous		771.5		309.6		322.5		794.0		1465		374.7
Second Pass	Unambiguous	-0.13	356.1	0.15	76.64	-0.38	124.8	-0.59	219.1	0.32	123.4	-0.23	197.4
	Ambiguous		354.5		77.73		119.5		210.8		130.5		193.6
Total Time	Unambiguous	-0.07	1033	0.47	297.0	-0.60	352.1	0.17	537.1	-0.64	572.8	0.24	458.5
	Ambiguous		1039		300.6		342.4		545.3		558.4		457.2
Probability of Fixation	Unambiguous	1.52	.93	-2.24 *	.38	2.72 **	.60	-0.96	.86	-0.01	.80	1.08	.82
	Ambiguous		.94		.34		.65		.86		.80		.85
Regressions Out	Unambiguous	0.50	.03	0.72	.15	-0.53	.18	-0.12	.33	-0.32	.78	0.04	.15
	Ambiguous		.03		.16		.16		.33		.77		.15
Regressions In	Unambiguous	0.18	.52	0.21	.26	0.10	.36	-0.46	.38	/	/	0.43	.45
	Ambiguous		.52		.26		.36		.36		/		.46

Table 10.11: Combined Analysis of Experiments 3 & 4 - Main Effects of Load. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if participants are under **Load** or **NoLoad**. Effects are also marked if **Marginal**.

<b>Load</b>	<b>Levels</b>	<b>Initial</b>		<b>Pre-Critical</b>		<b>Critical</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Pre-Crit + Crit</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	NoLoad	-2.21	770.8	-1.58	232.8	0.27	236.8	-1.43	353.6	0.86	377.6	-1.97 <sub>m</sub>	293.5
	Load	*	675.3		227.9		242.2		338.7		434.5		274.1
First Fixation	NoLoad	-0.42	219.2	-1.34	222.9	0.31	225.4	-0.50	24.5	0.50	239.3	-0.43	228.6
	Load		215.4		220.3		229.9		241.8		250.8		229.1
Regression Path	NoLoad	-2.13	812.2	-1.49	308.7	-1.01	331.0	-0.29	688.2	3.01	1295	-1.95 <sub>m</sub>	388.2
	Load	*	733.5		294.7		320.6		856.3	**	1683		362.6
Second Pass	NoLoad	2.24	286.6	-0.29	78.76	-1.42	111.64	0.21	212.6	-0.51	129.9	1.20	181.5
	Load	*	424.0		75.61		132.7		217.2		124.0		209.5
Total Time	NoLoad	-0.25	1014	-0.45	295.0	0.72	337.5	-1.29	548.5	0.31	538.4	-0.74	456.0
	Load		1059		302.6		357.1		534.0		592.8		459.7
Probability of Fixation	NoLoad	-1.14	.94	-2.04	.39	-1.94	.66	-2.96	.89	-1.44	.82	-2.66	.88
	Load		.94	*	.33	<sub>m</sub>	.59	**	.83		.78	**	.79
Regressions Out	NoLoad	1.25	.03	-1.17	.18	-2.35	.21	0.21	.32	2.02	.74	-1.11	.71
	Load		.03		.13	*	.14		.34	*	.81		.14
Regressions In	NoLoad	2.05	.47	-1.18	.28	0.80	.34	-0.04	.37	/	/	0.30	.45
	Load	*	.57		.24		.37		.37		/		.46

## 10.4 Discussion

In the introduction to this chapter we examined the findings which appeared to differ between the Load and No Load experiments. There appeared to be little variance in the trends of the main effects (although exact position and measure did vary), but critically the principal interaction of interest in the NoLoad experiment (the probability of fixating the critical region appearing modulated by Context and Ambiguity), was not significant in the Load experiment. Unlike in the Syntax chapter however, this was not a clear cut removal of all influence of Context and Ambiguity; instead it manifested as Context, Ambiguity and Frequency interaction in which Low frequency words were unaffected by Context or Ambiguity, but High frequency words showed evidence of being influenced (although not in an easily interpretable pattern).

We conducted a statistical meta-analysis to verify the reliability of the effects we were seeing. By adding Load as a factor into the analysis we were able to observe main effects of Load and importantly, interactions. As mentioned in the Syntax section, a main effect of Load is difficult to interpret as it indicates only a difference between the groups, which could be our Load manipulation, but might also be inherent differences between groups which we did not predict (such as the Load group having a faster average reading speed). This is unlikely, however, as participants were randomly assigned to groups. Regardless, interactions with Load are much less likely to be unpredictable variance as they are very specific and systematic.

The main effects of Context and Ambiguity in the combined analysis can be summarised thus: Context had no observable influence in early reading measures but later reading measures and regressions showed a pronounced negative influence of a Neutral context. It is likely that when a context is Supportive, less (or less extensive) re-reading is required. Ambiguity had no influence on reading measures but is implicated in directing fixations to the critical region. This is interesting as models of eye-movement control do not allow for pre-processing such that a word could be identified as ambiguous prior to fixation. Our secondary factors were Frequency (of the critical word) and because of the uncertainty of the effect's locus, Load. Frequency showed little effect but was apparent sporadically in the early reading measures for pre-critical region, implying that the frequency of the upcoming word was being identified in the para-fovea, and once in the total time spent in the wrapup

region, which implies readers found it harder to perform sentence integration and wrapup for low-frequency words. Load influenced the probability of fixation and the likelihood of regressions such that the critical regions received fewer fixations and regressions out.

The principal interactions were: A first pass interaction in the spillover region between Context, Ambiguity, and Load, and a probability of fixation interaction in the critical region between Context, Ambiguity, and Load. We will discuss these, and any supporting interactions, in turn.

In the Spillover region, the Context, Ambiguity, and Load interaction appears to show, for the non-Loaded participants, the pattern that Duffy et al. (1988) observed on the critical region: an elimination of the additional load of creating and maintaining two interpretations of an ambiguity, based on the presence of a preceding context. The difference here is that we are observing this on the spillover region rather than the critical region. However, assuming that this is just a normal delay in processing, it does allow us to fit our results with Miyake et al. (1994)'s Capacity Constrained Model. Miyake predicts that those with limited resources will struggle more to maintain multiple interpretations and will, when a context does not provide speedy selection, show greater influence of ambiguity than those with less limited resources (an increase in reading times for our Neutral / Ambiguous condition across Load states). When the context does provide support for the resolution of the ambiguity then High and Low span readers should behave similarly, and the ambiguity should be processed in almost the same way as a non-ambiguous word (with a slight elongation of reading times to account for multiple meaning access and subsequent selection). We do not, however observe this pattern. Instead we see a reduction of reading times for the Neutral / Ambiguous condition for those under Load (compared to those not under Load), to the point where there are no statistical differences between any of the conditions for those who are under Load. This implies that, rather than the differences predicted by Miyake et al. for High and Low span readers, our capacity constrained readers were not processing the ambiguities differently to the non-ambiguous controls at all, regardless of the presence of a supportive context. This runs against the predictions for High and Low span readers but may be due to experimental differences. Firstly, our readers were not pre-determinedly High or Low span, instead they were titrated down to a level of capacity constraint which may be lower than Low span readers would normally present. Secondly, the nature of reducing available resources by splitting

attention between two tasks is very different to the idea of simply having a naturally reduced overall capacity. The SR and SLIR theories would both predict this pattern of results (a functionally modular separation of language processes (post-interpretive language processes in the case of the SLIR) such that those under Load are unable to apply context assist with the processing of a lexical ambiguity). Neither however, would predict that ambiguities would be processed in the same way as unambiguous controls. It is possible that readers under load are more likely to make speedy selections based on frequency alone (rather than involving contextual support, or waiting to locate further disambiguating information); this interaction is observed in the spillover region, so speeded frequency based selection might have already occurred by this stage.

The probability of fixating the critical region is, as was discussed above, mediated by ambiguity and context in a very similar way to how Duffy et al. showed First Pass reading times to be influenced. This interaction disappeared under load and the three way interaction is significant. However, it is not the case that there is no influence of Context or Ambiguity when Loaded, as frequency also comes into play: There is a Context, Ambiguity, and Frequency interaction such that no influence of Context or Ambiguity is seen on Low frequency words, but High frequency words are affected. However, the pattern of the High frequency words does not replicate that seen in the NoLoad component of the three way context, Ambiguity, and Load interaction. This implies that, not only do participants treat Low and high frequency words differently, they do not treat High frequency words the same as a reader who is not under Load treats all words. We will return to this shortly. First, we will discuss the nature and implications of the contextual mitigation of ambiguity processing seen when participants are not under Load.

This interaction is observed in the probability of fixating the critical region. It shows a pattern of additional likelihood of fixation if the upcoming word is Ambiguous, with this additional likelihood being suppressed by the presence of a supportive context. This is important for two reasons: Firstly, it means that to some extent readers are able to identify ambiguous words in advance of fixations, and secondly, they are able to perform some kind of pre-processing which involves contextual information such that they feel able to skip the critical word in the same proportions as if it were unambiguous. Both of these assertions are problematic for models of eye-movement control (more so for EZ-Reader than SWIFT due to the latter's gradient processing system) which in general only allow familiarity

checks and very cursory pre-processing to occur in the para-fovea. This should not allow for the identification of words as ambiguous, nor should it allow for contextual influence at this stage. The observed interaction is given credence, however, by the fact that it follows the pattern seen for the NoLoad participants in the Context, Ambiguity, and Load interaction in First Pass reading times on the spillover region, and that this pattern is what is predicted and displayed by Duffy et al. (1988) and their re-ordered access model (albeit one region later on).

To return to the paired three way interaction of Context, Ambiguity, and Load, and Context, Ambiguity, and Frequency: The pattern we see for the High Frequency words under Load is that readers are able to recognise whether the next word is ambiguous or not, but are not able to use context to mediate the detrimental effects of accessing two meanings and selecting between them. The fact that the recognition of an upcoming ambiguity can only occur for high frequency words has two potential explanations. Firstly, that there is a resource cost associated with the mechanism by which they are identified, and that cost is inflated to untenable levels (while under heavy capacity constraints) for low frequency words which are generally harder to process. Secondly, it may be the case that, the bias that low frequency words are always hard to process and so are in general more likely to be fixated than high frequency words may be overriding the identification mechanism and causing a fixation regardless of ambiguity. This latter option may be true of Loaded readers as they have reduced capacity to handle harder low frequency words compared to those not under capacity constraints. It is also supported by the two way Load by Lexical Frequency interaction seen in the same region and measures (probability of fixating the critical region) which shows no difference in the fixation of Low frequency words across Load conditions.

The fact that context is not being used by Loaded readers to help disambiguate ambiguity is our main focus however, as it speaks to the SR vs SLIR discussion. Both theories predict a certain reduction in the use of context in ambiguity resolution, but while the SR theory indicates an overall reduction in contextual involvement under Load, the SLIR theory predicts that normal (interpretive) processing will continue to be facilitated by context even under Load. The interactions we have discussed indicate that Context is being brought to bear on ambiguity resolution when resources are not limited, but that it is not when participants are under Load. We now turn to the main effects, and interactions between



context and Load to help distinguish between the two theories. The main effects imply a general facilitation of processing when a supportive context is present. This general trend is indicated by the presence of main effects of context in second pass reading times and regressions in such that there are longer reading times and more regression in to each region when the context is Neutral. These main effects were also present in the Load only analysis, implying that contextual facilitation of normal processing is not affected by being under conditions of limited resources. There are two interactions between Load and Context, both in the probability of fixation measure, but neither of which are particularly trustworthy. Both have observations near ceiling, and while the one on the combined pre-critical and critical region represents the likelihood of fixating two separate regions, and is superseded in one of the regions by a three way interaction involving ambiguity, the other is in the wrapup region which brings its own biases of fixation. Nevertheless, the data within each interaction indicates contextual influence occurring when readers are under load. With both of those interactions being in dubious support, and no other Load by Context interactions observed in the experiment, it appears that contextual facilitation of normal processing continues unhindered by being under cognitive load. The data from this combined analysis therefore show support for the SLIR theory.

## **Chapter 11: Load, Context, and Lexical Ambiguity Resolution - Within Subjects (Experiment 5)**

### **11.1 Introduction**

In the previous meta-analysis we found support for the SLIR theory via three principal findings. The first was a Context, Ambiguity, and Load interaction in First Pass, spillover region, which for those not under Load mirrored Duffy et al. (1988)'s findings regarding balanced ambiguities and preceding contexts and supported the re-ordered access model, and for those under Load showed no use of context and a reversion to a frequency only method of processing of ambiguities.

The second was a Context, Ambiguity, and Load interaction in the probability of fixating the critical region which for those not under Load showed contextual influence, dependant on ambiguity, on the generation of saccades into the critical region. The pattern was such that when the context was neutral the probability of a saccade into the region was higher in the case of an ambiguity than a frequency matched unambiguous control, whereas when the context was supportive there was no difference predicated on ambiguity. The pattern altered under Load: there was no influence of either context or ambiguity on Low frequency words, but High frequency words showed evidence of readers identifying ambiguities in the para-fovea, but not using context to assist in pre-processing or as a guide to influence fixation behaviour. This revealed two separable mechanisms which eye-movement control models struggle to accommodate, the first is the identification of an upcoming ambiguity in the para-fovea (which requires pre-processing to a greater extent than most models allow), and the second is contextual influence on that pre-processing.

The third was a lack of reliable Context by Load interactions, or evidence to support the assertion that Load was having a negative influence on the use of context in normal processing (that not involving a particular challenge such as lexical ambiguity).

In this, the final experiment in this series on Lexical ambiguity, we address the main negative point regarding our experimental design. Up to this point we have manipulated two factors (Context and Ambiguity) within subjects, and a third (Cognitive Load) between subjects. As has been mentioned, main effects of Load are not wholly attributable to our Load manipulation as other, unplanned and unpredicted factors may have varied between

our NoLoad and Load groups. In this experiment we will be manipulating Load as a within subjects factor. This means that any main effects or interactions with Load can be directly attributable to a reduction in the reader's available resources. To avoid having 2x2x2 design with Context, Ambiguity, and Load (which would require much more power to show statistical effects than a 2x2) we removed the ambiguity variable, choosing to have only ambiguous words in the experimental items. We did this because we are most interested in replicating the critical interactions we saw in the First Pass of the Spillover region, and the Probability of Fixating the Critical region. Both of these interactions are driven by variance in the ambiguous condition across load, while the unambiguous condition remains static. We reasoned that the experiments discussed above have provided insight into the differences in processing for ambiguous and unambiguous words under load, and can provide a point of comparison for that variable should it be necessary, but that we could reduce the power requirements by removing the Unambiguous condition, and still retain the experimental variables required to test if the interactions replicated in a within subjects design.

## 11.2 Methods

### *Participants*

37 participants took part in the study. The requirements were the same as the previous two experiments, with the addition that participants had not taken part in either of those experiments.

Total Exclusions based on poor comprehension, tracking, or digit recall = 1

Comprehension Question Range (of included participants) = 83% - 97%, Mean = 92%.

Digit Span Range (of included participants) = 5 – 9, Mean = 6.

### *Stimuli*

In this experiment we removed the Ambiguity manipulation, choosing to focus on Ambiguous words only and forego the unambiguous controls. This meant that the counterbalancing, which had kept participants from seeing the same balanced ambiguous word twice, could no longer be maintained. Consequentially, no new items were created for this experiment, but the full item list (see Appendix 2), minus items containing unambiguous controls was displayed to each participant.

A set of example stimuli is given in 0 - 11.2.d.

11.2.a. The woman was unhappy with the way things were going.  
It was no good, the appendix would have to be removed.  
(*Neutral* NoLoad)

11.2.b. The woman was complaining of abdominal pain.  
It was no good, the appendix would have to be removed.  
(*Supportive* NoLoad)

11.2.c. The woman was unhappy with the way things were going.  
It was no good, the appendix would have to be removed.  
(*Neutral* NoLoad)

11.2.d. The woman was complaining of abdominal pain.  
It was no good, the appendix would have to be removed.  
(*Supportive* NoLoad)

In order to minimise any effects of lexical repetition (of the balanced ambiguous words) and reduce the potential interference of task switching (see Literature Review sections 1.2.4.1 and 1.4) a mini-blocking pseudo randomisation approach was adopted. Blocking was four items deep and alternated between NoLoad and Load by block. Within a block, the first item was a randomly selected filler, the second was a randomly selected experimental item, and the third and fourth were randomly selected fillers. This meant that the state of Loading did not change on an experimental item, and that experimental items were always a minimum of four trials apart. The selection of experimental items within the blocking system was randomised so that while it was possible for a particular critical noun to be seen twice within four trials, there was no structured or conditional bias dictating how this would occur. A second experiment (reported in Section 4: Chapter 13) was run concurrently with this one and experimental items for that experiment took the place of Filler 2 in the example of the blocking system that is given in 11.2.e:

#### 11.2.e. Mini-Blocking Schematic

NoLoad Block begins  
Filler 1  
Experimental Item 1  
Filler 2  
Filler 3  
NoLoad Block ends

Load Block begins  
Filler 4  
Experimental Item 2  
Filler 5  
Filler 6  
Load Block ends

#### *Procedure*

The same digit span pre-test and pre-rest procedures were used to tailor the level of difficulty to each participant. It was explained that the experiment would only have a Digit component for half of the items. For the blocks without Loading, the presentation and procedure was identical to that of experiment 3, for the blocks with Loading, the presentation and procedure was identical to that of experiment 4.

#### *Data Analysis*

The procedures used to analyse experiment 5 were identical to those used in experiment 3 and experiment 4, with the exception of the addition of a Repetition predictor (and the removal of the Ambiguity predictor). The repetition of the critical nouns in this experiment is likely to have influenced reading behaviour (Anderson & Holcomb, 2005; Scarborough, Cortese, & Scarborough, 1977; Scarborough, Gerard, & Cortese, 1979), and so it was necessary to factor this into the analysis. A two level predictor was created, which distinguished between whether, for a particular participant, the critical noun was novel (FirstMention), or had been seen before (SecondMention). We include the Repetition and Frequency as factors in the analysis because we are aware that they may have a systematic influence on our results. We want to partial out this variance so that only the true picture of the way that our principal manipulations interact remains.

### 11.3 Results

*/It was no good /the /appendix /would have/ to be removed./*  
*/        2        / 3 /        4        /        5        /        6        /*

1 = Context, 2 = Initial, 3 = Pre-Critical, 4 = Critical, 5 = Spillover, 6 = Wrapup

#### Overview of the Critical Findings:

The intent of this experiment was to replicate the major significant finding of the between subjects experiments, which was that, compared to those who were not-loaded, participants under load showed less interaction between contextual information and lexical ambiguity when planning whether or not to fixate the critical word. Although we do find some evidence of Context and Load interacting, we do not replicate this result (see Figure 11.1).

We do find some evidence of Context and Load interacting in this experiment: Firstly in the Spillover region for First Pass (see Figure 11.2), and secondly in the total time spent on the Pre-Critical region (see Figure 11.3). Interestingly, in the previous combined analysis we observed interactions between Context, Ambiguity, and Load in both these regions / measures, but while the Total Time interaction replicates the pattern seen for the ambiguous words in the combined analysis, the first pass interaction does not.

#### Context by Load Interactions:

Probability of Fixating the Critical Noun (non-significant):

In Probability of Fixation for the previous experiments in this chapter we observed an interesting three way interaction between Context, Ambiguity and Cognitive Load in which we observed a Context and Ambiguity interaction which disappeared under Load. When not Loaded we saw a higher likelihood of fixating the Critical Region if the Critical Word was Ambiguous rather than Unambiguous (but only if the Context was Neutral). This was a statistically significant interaction. When Loaded we observed a main effect of Ambiguity but no interaction.

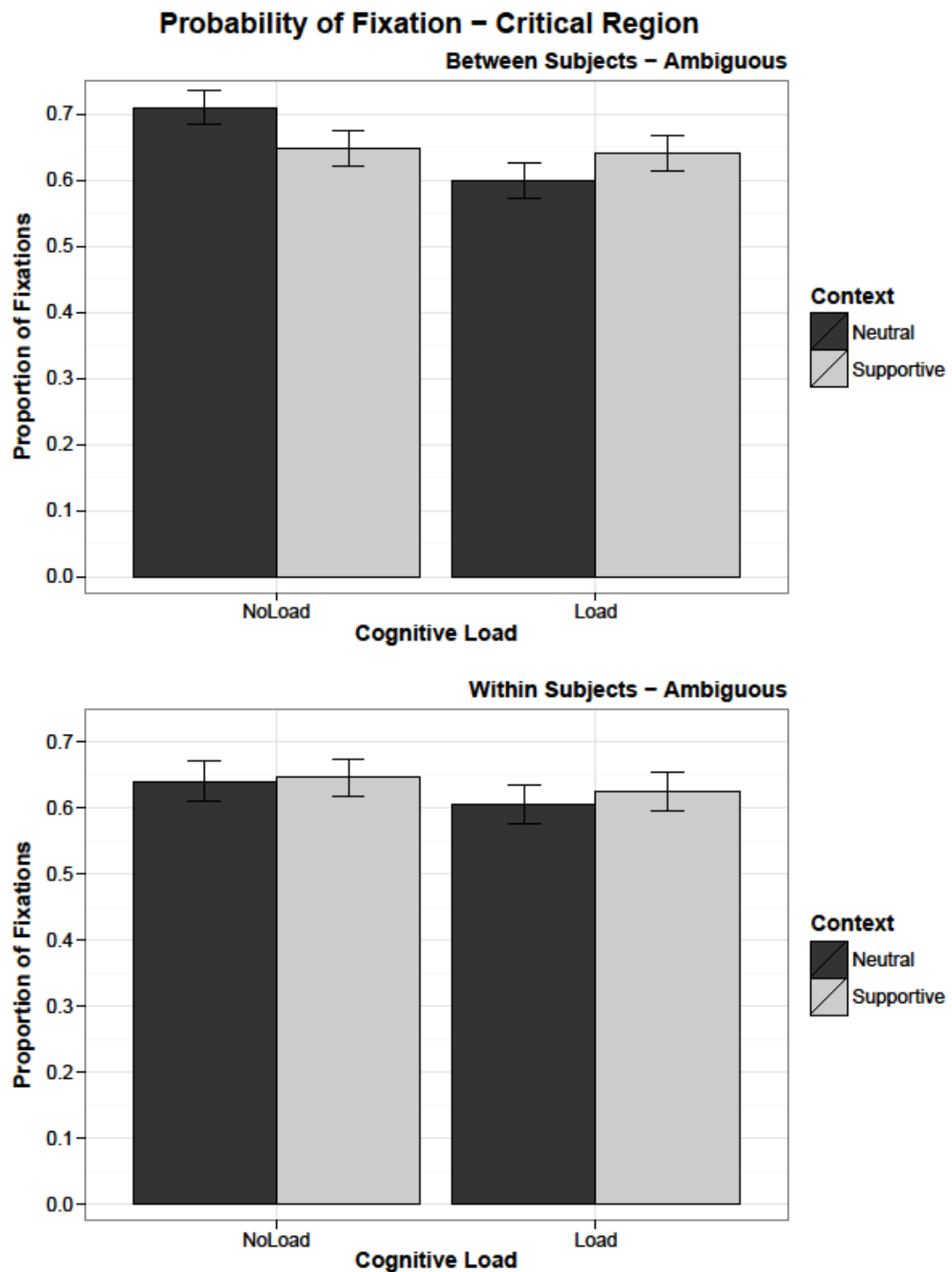


Figure 11.1: Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Context and Cognitive Load Interactions for Between Subjects and Within Subject designs. Ambiguous words only.

Looking only at the observations in which the critical word was Ambiguous we observed a significant Context by Cognitive Load interaction. When not under Cognitive Load we saw that participants were more likely to fixate the Critical Region when the Context was



Neutral than when it was Supportive. When under Load, we observed no statistical difference in fixation likelihood. It appeared that, when not loaded, fixation probability followed the pattern that Duffy et al. (1988) would expect to see in reading times: A penalty for encountering an Ambiguous word that is not disambiguated by the Context. But when Loaded, participants were less able to make use of a Supportive Context and were equally likely to fixate the Critical Word, regardless of Context.

Within subjects we observe no Context by Cognitive Load interaction in the Critical Region for Probability of fixation ( $p > 0.1$ ) (Figure 11.1, Panel 2). There are no significant or marginal pairwise effects and numerically the conditions varied little (NoLoad / Neutral = 63, NoLoad / Supportive = 64, Load / Neutral = 60, Load / Supportive = 62). We were unable to replicate the finding observed in the between subjects design (Figure 11.1, Panel 1). This will be discussed in more depth in section 11.4 and Chapter 12.

#### First Pass for the Spillover Region:

In First Pass we observe two relevant interactions, both in the Spillover Region. The first is a Context by Cognitive Load Interaction ( $p = 0.012 *$ ) (see Figure 11.2, Panel 2). When not under Cognitive Load we observe approximately equal First Pass fixation durations across Context (Neutral = 343ms vs Supportive = 340ms,  $p > 0.1$ ). Whereas, when under Cognitive Load, participants displayed longer First Pass duration when the Context was Neutral, as compared to when it was Supportive (350ms vs 302ms,  $p < 0.001 ***$ ). When the Context is Supportive we see significantly longer First Pass fixation durations across Cognitive Load (NoLoad = 340ms vs Load = 302ms,  $p < 0.001 ***$ ), whereas when the Context is Neutral we see no significant difference across Cognitive Load (NoLoad = 343ms vs Load = 350ms,  $p > 0.1$ ). The data indicate no influence of context on the processing of ambiguous words while not under Load, but speeded processing of ambiguous words when a supportive context is present while under Load. This is not the same pattern as was observed in the between subjects analysis, in which there were no significant pairwise interactions (see Figure 11.2, Panel 1). Between the two experiments we appear to see a suppression of the Load / Supportive condition.

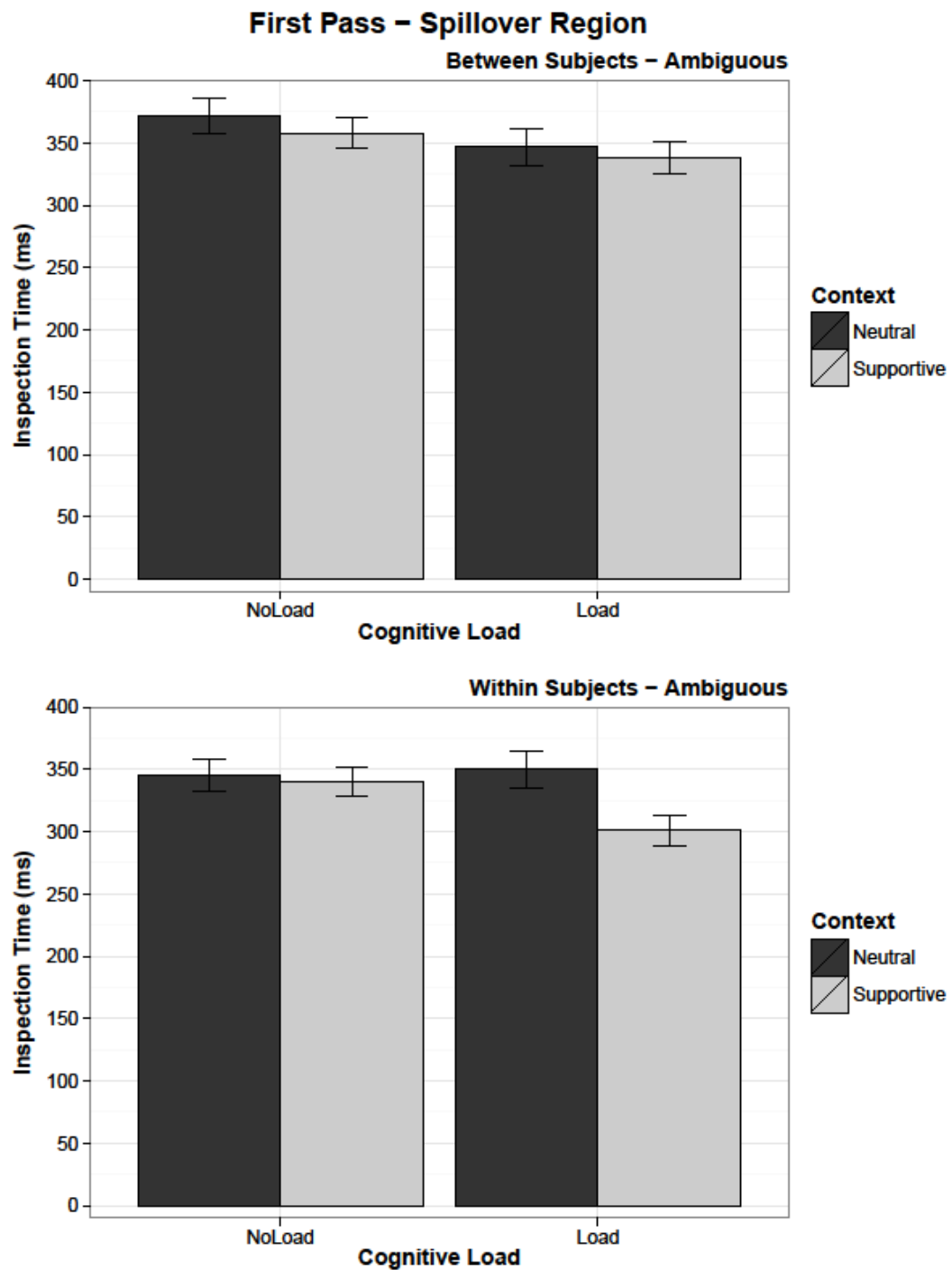


Figure 11.2: Mean inspection time for First Pass in the Spillover Region. Error Bars display Standard Error. Context and Cognitive Load Interactions for Between Subjects and Within Subject designs. Ambiguous words only.

#### Total Time spent on the Pre-Critical Region:

In the Pre-Critical region we observe a Context by Cognitive Load interaction ( $p = 0.021$  \*) (see Figure 11.3, Panel 2). When the participant is not under Cognitive Load we see no difference across Context (Neutral = 292ms vs Supportive = 303ms,  $p > 0.1$ ), however, under Load we observe a significant difference across Context (Neutral = 319ms vs Supportive = 273ms,  $p = 0.003$  \*\*). Across Load, we observe no significant effect if the Context is Neutral (NoLoad = 292ms vs Load = 319ms,  $p > 0.1$ ), but a significant difference when the Context is Supportive (NoLoad = 303ms vs Load = 273ms,  $p = 0.024$  \*). It appears then, when under Load, there is both a numerical slow down when the Context is Neutral (compared to the NoLoad / Neutral condition), and a significant speed up when the Context is Supportive. Why this should be is unclear, but it replicates the pattern for this measure in this region which was observed in the combined analysis (see Figure 11.3, Panel 1).

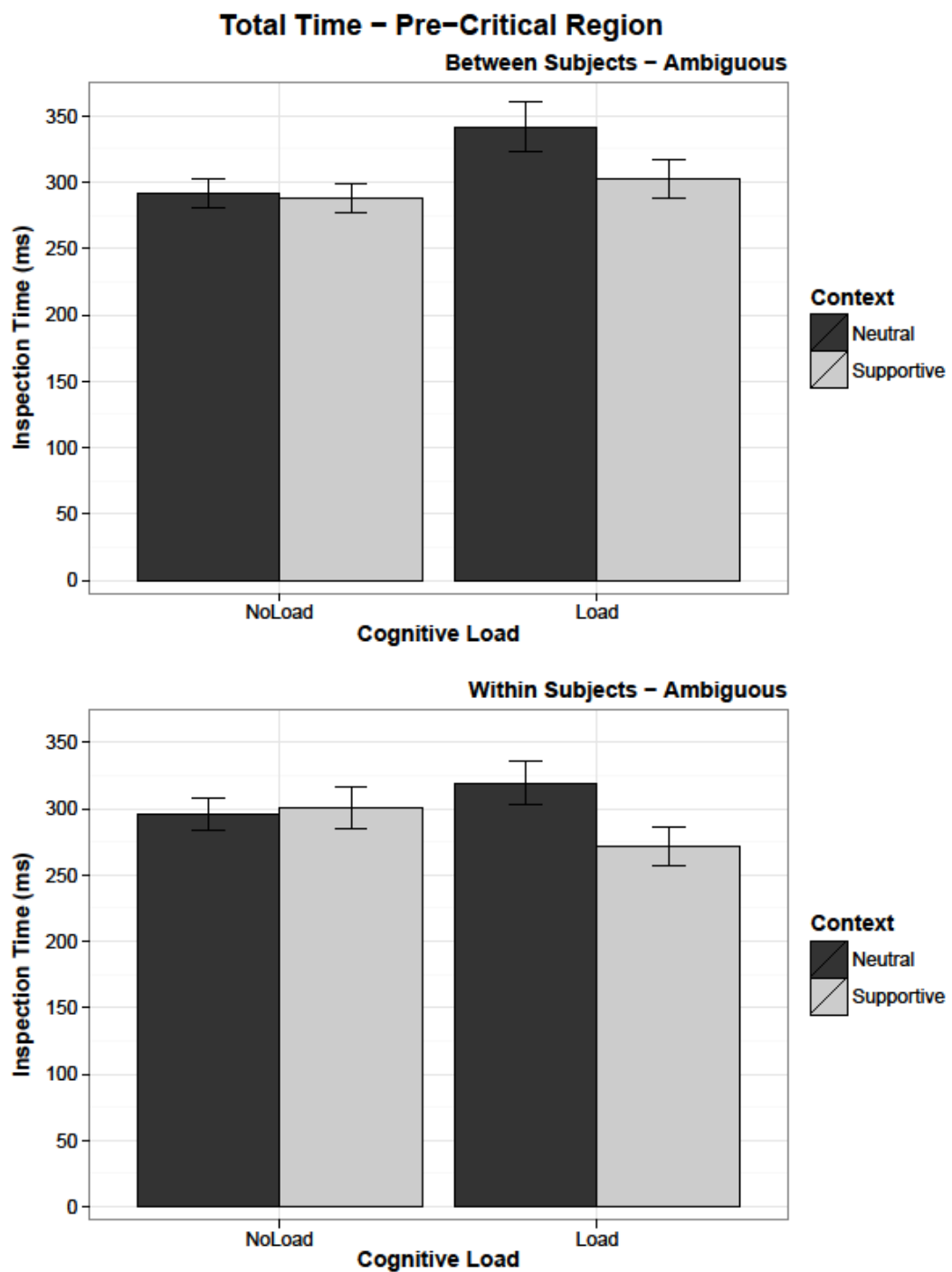


Figure 11.3: Mean inspection time for Total Time on the Pre-Critical Region. Error Bars display Standard Error. Context and Cognitive Load Interactions for Between Subjects and Within Subject designs. Ambiguous words only.

### Other Interactions:

In the sections below we will discuss any relevant interactions, focusing on those which involve our principal manipulations: Context and Load. It is worth noting that we observe many interactions involving Repetition and Lexical Frequency and, while interesting from a Lexical Processing perspective, these will not be discussed in this thesis unless they have a particular bearing on the interpretation of the principal manipulations.

Table 11.1 lists the interactions involving Frequency and / or Repetition which we will not discuss in depth.

Table 11.1: Experiment 5 - Interactions involving Frequency and / or Repetition which are not reported in the Results section.

Factors	Measure	Region
Context, Freq	Total Time Regression Path	Critical, Wrapup Wrapup
Load, Freq	Probability of Fixation	Pre-Crit+Crit
Load, Rep	Second Pass Probability of Fixation Regressions In	Pre-Crit+Crit, Crit, Wrapup Spillover Initial
Context, Rep	Regression Path	Spillover
Freq, Rep	First Fixation	Initial, Pre-Crit+Crit
Context, Load, Freq	Regressions Out	Critical
Context, Freq, Rep	First Fixation Regressions Out	Initial Critical
Load, Freq, Rep	First Pass Regression Path First Fixation	Pre-Crit Pre-Crit Pre-Crit, Critical

### Context by Repetition Interaction:

#### First Pass of the Spillover Region:

We see a Context by Repetition interaction ( $p = 0.034^*$ ) in the spillover region for first pass (see Figure 11.4). When the Context is Neutral we observe no difference in FP fixation durations, regardless of whether the Critical Word is novel (First Mention) or has been seen already (Second Mention) by the participant (346ms vs 347ms,  $p > 0.1$ ). When the Context is Supportive, we observe a marginal difference across Repetition (First Mention = 332ms, vs Second Mention = 312ms,  $p = 0.072^m$ ). We also observe a significant difference across Context when the Critical Word is a Second Mention (Neutral = 347ms vs Supportive = 312ms,  $p = 0.005^{**}$ ). It appears then, that there is a greater facilitation effect for Second Mention if the Context is Supportive of a particular interpretation of the ambiguous Critical Word. We report this as it indicates that higher level factors like context can quickly (although not immediately) influence lexical repetition / priming effects.

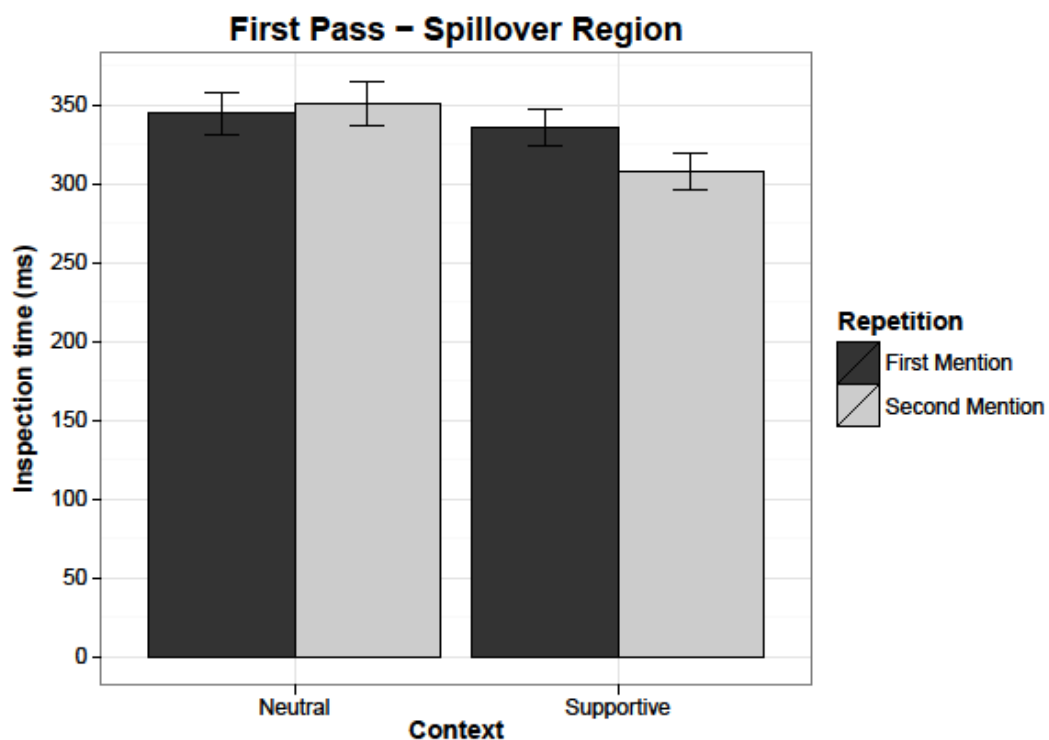


Figure 11.4: Experiment 5 – Mean inspection time for First Pass in the Spillover Region. Error Bars display Standard Error. Context by Repetition Interaction.

### **Summary of Main Effects:**

The main effects are reported in the tables below. Table 11.2 reports Context effects, while Table 11.3 reports Ambiguity effects.

#### **Context:**

Table 11.2 shows the normal pattern of contextual facilitation (shorter fixations, reading times, and re-reading times, and a lower proportion of fixations and regressions in and out, if the context is supportive). Interestingly, the Critical noun appears mostly unaffected, with the majority of context effects occurring on the preceding or spillover regions. The one exception to the general trend is a tendency for a supportive context to cause a longer regression path from the initial region. This makes sense as those regressing from that region have more propositional content to process in their re-reading, if the context is supportive.

#### **Load:**

Table 11.3 shows a far more mixed pattern of Loading than we have seen in the between subjects experiments. This, perhaps, implies that our load manipulation here was not as potent as in the previously experiments. This will be discussed more in the next section (11.4) and in Chapter 12). Overall the trend is towards shorter fixations and reading times when not under Load, compared to when Loaded, although there are exceptions to this on the critical region for early reading measures (FP, FF). In those cases the effect patterns in the opposite direction. Critically for the probability of fixation interaction which we sought to replicate, in this experiment it appears that the usual increase in skipping while under Load is not present, and the reverse pattern is observed.

#### **Frequency:**

There is a noteworthy frequency effect on the probability of fixating the Pre-Critical + Critical region, such that a low frequency word receives a higher proportion of fixations.

#### **Repetition:**

There are main effects of repetition on re-reading times (SP) and the proportion of regressions (RI) for all regions such that a FirstMention word generates longer re-reading times and a higher proportion of regressions than a SecondMention word.

Table 11.2: Experiment 5 - Main Effects of Context. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if Context is Neutral or Supportive. Effects are also marked if Marginal.

<b>Context</b>	<b>Levels</b>	<b>Initial</b>		<b>Pre-Critical</b>		<b>Critical</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Pre-Crit + Crit</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Neutral	0.56	714.5	-2.38	234.8	-0.28	234.3	-2.50	344.7	0.31	393.2	-2.68	277.8
	Supportive		716.7	*	212.9		236.7	*	320.8		402.5	**	259.6
First Fixation	Neutral	-2.87	216.7	-2.23	225.6	-0.84	226.9	-1.30	245.5	-0.55	241.4	-1.97	229.4
	Supportive	**	208.9	*	208.9		224.9		234.8		238.8	*	222.8
Regression Path	Neutral	1.98	770.2	-0.15	311.8	-0.31	331.6	-2.47	741.4	-1.02	1815	-1.10	375.5
	Supportive	*	803.1		332.0		346.8	*	662.5		1689		377.7
Second Pass	Neutral	-2.53	384.5	-0.60	86.13	-1.19	118.4	-2.36	259.5	-0.87	180.2	-1.04	200.7
	Supportive	*	312.6		78.76		107.7	*	211.0		151.8		184.7
Total Time	Neutral	-1.88	1076	-1.90	302.1	-1.29	338.2	-3.45	593.0	-0.20	624.4	-3.49	465.2
	Supportive	<sup>m</sup>	1001	<sup>m</sup>	283.1		335.8	***	524.7		586.0	***	434.7
Probability of Fixation	Neutral	-0.41	.96	-2.08	.37	0.44	.62	-0.09	.89	0.69	.79	-0.80	.85
	Supportive		.95	*	.31		.63		.89		.82		.84
Regressions Out	Neutral	1.59	.03	0.98	.15	-0.27	.20	-2.20	.34	-0.86	.79	1.03	.15
	Supportive		.05		.17		.18	*	.27		.77		.17
Regressions In	Neutral	-2.62	.54	-1.30	.30	-2.44	.37	-0.45	.39	/	/	-2.98	.49
	Supportive	**	.47		.27	*	.31		.38		/	**	.41



Table 11.3: Experiment 5 - Main Effects of Load. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if participants are under **Load** or **NoLoad**. Effects are also marked if **Marginal**.

<b>Load</b>	<b>Levels</b>	<b>Initial</b>		<b>PreCritical</b>		<b>Critical</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>PreCrit Crit</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	NoLoad	2.29	726.3	1.49	228.7	-2.62	227.9	2.79	341.4	-0.76	395.6	0.59	269.9
	Load	*	705.6		219.0	**	243.1		324.1		400.1		267.5
First Fixation	NoLoad	-1.90	205.2	2.12	225.0	-3.17	217.5	0.17	239.0	-3.04	229.3	-0.68	224.3
	Load	<sup>m</sup>	220.5	*	209.6	**	234.3		241.3	**	250.9		227.9
Regression Path	NoLoad	1.77	771.7	-0.08	296.2	-0.30	336.7	0.15	651.9	2.39	1764	0.75	368.1
	Load		801.6		347.6		341.7		752.0	*	1739		385.1
Second Pass	NoLoad	-2.31	301.9	-0.01	82.42	-0.45	108.8	0.55	239.4	0.17	166.3	-0.53	186.3
	Load	*	395.3		82.46		117.3		231.0		165.7		199.0
Total Time	NoLoad	-0.20	995.6	0.94	297.0	-0.01	329.7	4.12	573.3	1.74	601.8	1.78	451.0
	Load		1081		288.2		344.2	***	544.3		608.6		448.9
Probability of Fixation	NoLoad	3.83	.98	3.68	.38	1.08	.64	3.67	.92	1.71	.82	4.22	.89
	Load	***	.93	***	.29		.61	***	.86		.79	***	.80
Regressions Out	NoLoad	-0.08	.04	0.03	.17	0.83	.19	-0.36	.30	3.30	.82	0.94	.17
	Load		.04		.15		.19		.31	***	.74		.15
Regressions In	NoLoad	0.21	.51	1.50	.30	0.77	.34	3.75	.43	/	/	1.86	.47
	Load		.50		.27		.34	***	.34		/	<sup>m</sup>	.43

## 11.4 Discussion

This experiment was conducted with the view of replicating the main findings of the between subjects meta-analysis were presented previously. Here we will discuss the overall findings of the experiment and how those findings compare to the principal findings of the between subject analysis.

Firstly, it is worth noting that this experiment does not speak particularly well to the SR, SLIR debate which has driven our previous discussions. This is because the two theories predict the same pattern of reading behaviour while processing challenges such as lexical ambiguity, and only differ in their predictions of contextual use in normal processing situations. In this experiment we have only lexical ambiguities, with no unambiguous controls to represent normal processing. Nevertheless the main effects indicate that context is having an effect on normal processing in the early reading measures (FP, FF, RP). In the majority of cases we observe longer reading times when the context is neutral compared to when it is supportive. We see less of an influence of context on the later reading measures (SP, TT) which implies that any influence it is having happens early on. Contrary to this, however, Context is also implicated in a higher proportion of regressions out (after the critical word) and regressions in to the earlier regions. Load is seen to be having an effect such that, under Load, fewer fixations are made over all, but those fixations that are made, are in general, longer.

Overall, we do see mixed evidence of Contextual influence of ambiguity resolution being mediated by Load. We see two interactions worthy of discussion.

Firstly, a Context by Load interaction in the Spillover region for First Pass. This is not the pattern we observed in the previous experiment: The previous experiment showed no differences across Load or Context when the critical word was Ambiguous. This experiment shows a comparative reduction in the time Loaded readers spent on the region when the context was Supportive (making it a shorter fixation in supportive contexts than in neutral ones). No other differences are observed. This is not the pattern we would expect to see as it indicates that there is no observed contextual facilitation for those not under load, but a positive effect of a supportive context when capacity is reduced. Three possibilities appear plausible: Firstly, readers under load may be equally able to utilise context clues, and are

more inclined to do so when they are under pressure to get to the end of the sentence and offload the digits. This goes against the general findings of the previous experiments reported in this thesis, which have indicated that contextual influence is partially, at least, reduced by the application of Load. Secondly, it may be that this represents a spillover effect in which we would expect to see residual processing from the previous region spilling over into this one. This explanation is dubious, however, because a). the pattern we are seeing doesn't correlate with the anticipated pattern of the previous region, and b). the previous region shows no signs of an interaction between these two factors. Finally, we observe a Context by Repetition interaction, and a marginal Context, Load, and Repetition interaction in the same region for the same measure, and it is possible that the involvement of Repetition is confusing the results.

We also see an equally confusing Context by Load interaction in the Total Time measure for the Pre-Critical region. This one appears to show participants not under Load processing the ambiguity the same regardless of context, whereas those under Load seem to find it much harder under a neutral context. This pattern is identical to that observed in the between subjects analysis. There are two possible interpretations of this: Firstly, it is not the case the supportive context is facilitating Loaded participants (as the reading times for that condition are the same as those for both contexts in the NoLoad condition), rather there is additional difficulty occurring as a result of a Neutral context. This implies that readers under load are sensitive to context, but that who are not under Load are not using context to facilitate processing. This goes against our own previous observations, and those established in the literature. Secondly, it may be that this interaction, being as it is in Total Time, represents a combination of two non-opposing effects from the First pass, and Second Pass reading measures. We see evidence of this in the trends for each measure but neither is even marginal for main effects or interactions.

The results of this experiment do not neatly fit with either the general findings of the preceding experiments or the established literature. We will now discuss one final critical finding of the between subjects experiments (the Context, Ambiguity, and Load interaction in the probability of fixating the critical region), before returning to consider why the experiments may be so inconsistent.

Taking just the ambiguous conditions from the between subjects experiments we can see that readers who are not under load are less likely to fixate the critical region if the context

has been supportive compared to neutral. We interpreted this in light of Duffy et al.'s considerations about the use of context to support faster selection between accessed meanings of an ambiguity. In this case, we posit that readers not under load were able to make use of context to reduce the overall difficulty of processing the ambiguity (including selection of a specific meaning) to such a level that if the need to directly fixate the word was reduced. When loaded, however, this facilitation was not present which indicated that a reduction in resources inhibited the ability to apply context, either in general or in particular to processing occurring in the para-fovea. In this within subjects experiment we see absolutely no indication of the facilitation effect when the readers are not loaded, and the same lack of variance across context when they are.

There are two main methodological reasons why the results of the between subjects and within subjects experiments might differ so much. Firstly, and most obviously, in the between subjects design we had two different groups of people undertaking the experiment separately. We have no reason to believe that one group should perform differently to another; both groups were drawn from the same population and conformed to the restrictions (age, native-ness, etc.). It is possible however, that some unpredicted variance was introduced due to the use of separate groups. Secondly, the tasks demands of each experiment were different. In the between subjects design participants were required to divide their attention between two tasks (by maintaining the digits while reading sentences) and to switch, on a within trial basis, between reading and digit recall. In the within subjects design participants had to perform the same attentional divide and within trial task-switch, but also had to switch between loading states every four trials. This may have increased the overall experimental difficulty compared to the NoLoad group, by giving them an additional task to keep track of (see Literature Review, sections 1.2.4.1 and 1.4). It may have also reduced the overall task demands compared to the between subjects Load group because the within subjects participants were, comparatively, only loaded half of the time. This differentiation might have put the overall task demands applied to the within subjects readers as somewhere in between those applied to the NoLoad and Load groups. As mentioned in the methodology of this chapter, steps were taken to reduce these differences (Load state switched were clearly marked, proceeded only when participants were ready, and occurred only on fillers rather than experimental items), but inherently the experimental designs do remain different.

## **Chapter 12: End of Lexical Section - Summary and Discussion**

In this section we have reported two experiments, a between subjects meta-analysis, and a within subjects experiment. The intention was to examine the effects of reducing the availability of cognitive resources on how context was used in the processing of lexical ambiguities. The first experiment provided a baseline under which cognitive resources were left as normal, the second experiment placed readers under a cognitive memory Load, and the meta-analysis compared the two. The third experiment was conducted to provide within subjects verification of differences found between the two previous experiments.

Seeking to replicate the findings of Duffy et al. (1988) in our lab we used preceding contexts and balanced lexical ambiguities. Duffy et al.'s re-ordered access model predicted slowdown on the critical word for balanced ambiguities compared to non-ambiguous controls due to simultaneous access and elaboration of each meaning of the ambiguous word. This slowdown could be mitigated by the presence of context which strongly supported one of the meanings of the ambiguous word. In this case, both meanings are still accessed, but selection occurs very quickly as the activation level of one of the meanings far exceeds that of the other. Therefore by manipulating the supportiveness of the context, and whether the critical word was ambiguous or not, we created a 2x2 design crossing context and ambiguity, and predicted an interaction between the two on the critical word.

Miyake et al. (1994) proposed a capacity constrained model of lexical ambiguity resolution which predicted that High and Low span readers would differ in their processing of ambiguities based on the availability of resources. They predicted that, for balanced ambiguities, when there is a supportive context present there will be no difference between span groups (as the additional information from the context assists in very early selection of a particular meaning, and as such multiple elaborations are not maintained for long), whereas when a context is not supportive differences will be seen between the two groups. The Low span group, who are limited on resources, will struggle to maintain multiple interpretations for long and will show slowdown in the regions immediately after the ambiguity, and might eventually drop one interpretation in favour of the most likely (based on minor frequency differences). We have seen in the previous Syntax section that readers under an extrinsic memory load do not behave quite like those with Low working

memory spans, but these predictions still provide a rough guide to how we might expect Loaded readers to perform.

Above and beyond the level of Miyake et al.'s predictions are the two competing theories of resource sharing that drove the investigations in the Syntax section. With regards to lexical ambiguity, the SR and the SLIR theories offer similar predictions as they did for syntactic ambiguity. Both predict reduction of contextual involvement in the processing of ambiguities as resources are depleted, the SR because the overall resource pool is depleted and becomes more modular, the SLIR because the processing of ambiguities falls within the purview of post-interpretive processing, which draws from the same resource pool as higher level contextual processing (and the same modularity applies). The SR and SLIR theories differ, however, on their predictions of contextual influence on normal processing under conditions of resource depletion. The SR theory, with its single resource pool for all processes, predicts that contextual facilitation of normal (non-ambiguous) processing should be reduced under Load. Under the SLIR theory on the other, normal language processing is interpretive (rather than post-interpretive) and as such draws from a different resource pool to higher level contextual processing. A reduction of one pool, it is argued, should not affect processes which draw from another (and interpretive processes can still accept input from post-interpretive processes, regardless of their level of resource depletion).

#### *Summary of Findings:*

In the first experiment we saw contextual facilitation of normal processing, but only in re-reading times and the likelihood of regressing in to regions previously fixated. We saw main effects of Ambiguity but they appears to influence the likelihood of fixating the critical region, rather than reading times on ambiguous words as Duffy et al. would predict. We were also unable to replicate Duffy et al.'s Context by Ambiguity interaction on the Critical region in First Pass. We reasoned that this may be due to visual differences in the presentation of the context (Duffy et al.'s was either before or after and shifted the position of the critical word accordingly, whereas ours was always in the same positional but the content changed), or measurement differences (Duffy et al. used gaze duration with a variance of 4 characters back in cases of no fixations, whereas we used the more regular First Pass). However, neither of these explanations proved satisfactory and the more likely cause of the difference is that we saw a Context by Ambiguity interaction in the probability

of fixating the region which Duffy et al. saw their interaction in. The interaction was such that there was a higher probability of fixating the critical word if it was ambiguous and the context was not supportive of a particular interpretation of the ambiguity. When the context was supportive, the ambiguity caused no additional increase in the likelihood of fixating the critical region. This suggested that some greater level of contextually mediated pre-processing was occurring prior to fixation, which may have reduced the differences present on the critical word itself.

In the Loaded version of this experiment we see a similar trend in the main effects of Context: Regression measures and later reading measures indicate more and more thorough re-readings when the context is neutral. We see very few main effects of ambiguity, but these are apparent around the probability of fixating the pre-critical and critical regions. The interaction between Context and Ambiguity (Probability of fixating the critical region) that we saw in the NoLoad experiment was not apparent here. Instead a three way interaction between Context, Ambiguity, and Lexical Frequency such that low frequency words are unaffected by either context or ambiguity, and high frequency words are, but display a different pattern of influence to that which we see in the NoLoad experiment.

The statistical meta-analysis of the two experiments combined showed more similarities than differences in the main effects. Context effects still occurred in later reading measures and proportion of regressions in (longer reading times and more regression in in the context was neutral), and ambiguity appeared to only affect the probability of fixating the pre-critical and critical regions. Main effects of Load (as much as we are cautious with their interpretation) showed participants reading faster, less thoroughly and with fewer mid-sentence regressions when Loaded.

We observed two interactions worth discussing, both three way interactions involving Context, Ambiguity, and Load, one in the spillover Region for the First Pass measure, and one in the probability of fixating the critical region. The first appeared to show the pattern that Duffy et al. predicted (only one region downstream), where participants appeared to encounter more difficulty with ambiguous words if the context was not supportive than if it was, whereas unambiguous words saw no contextual facilitation. This interaction disappeared under Load (all conditions were equal). We took this as evidence to support the re-ordered access model, but not Miyake et al.'s capacity constrained model. Instead

we took a step back and compared the data to the predictions of the SL and SLIR theories. Both theories would predict a lack of contextual influence on ambiguity processing when under Load (which we see), but have any predictions relating to the fact that the ambiguities were being processed the same as non-ambiguous controls. We consider that by the time we see this interaction (in the Spillover region) those under Load have simply selected the most likely interpretation (based on frequency) and are therefore no longer maintaining two interpretations.

The second interaction (Context, Ambiguity, and Load; First Pass; Probability of Fixation; Critical Region) combined with a Context, Ambiguity, and Frequency interaction in the same region indicated that, when not under Load, readers were able to partially process the upcoming critical word and not only identify it as ambiguous, but also make use of contextual information reduce the potential difficulty of it to a level at which they felt no greater need to fully fixate it than an unambiguous control. Readers under Load, however, were only able to identify the upcoming word as ambiguous (and were only able to do that when it was easily accessible due to being of high frequency). As mentioned in previous discussions, this interpretation is contentious due to the restrictions on para-foveal processing assumed in current eye-movement control models.

At this stage, our data support the SLIR theory, which predicts contextual influence on normal processing being largely unaffected by a reduction in available resources, but contextual influence on ambiguity resolution being reduced / eliminated.

The final experiment in this series used the same materials, but adopted a mini-blocking approach using only ambiguous words, and a within subjects design to verify the two interactions mentioned above within subjects. Unfortunately, the Spillover region interaction did not replicate (with a different pattern showing what appears to be contextual facilitation in the processing of ambiguities in the Load condition (but not the NoLoad condition)), and neither did the probability of fixation interaction (with there being no difference between conditions at all). The main effects also showed a different pattern with context effects occurring in the early reading measures such that longer reading times were witnessed when the context was neutral. The main effects of Load showed the same pattern as before.



It seems then that we have support for the SLIR theory across both the between and within subjects experiments such that normal processing of language appears unaffected by the reduction of available resources. However, the SR and SLIR theories both predict a reduction in the influence of context on ambiguity resolution which we see in the between subjects design, but not in the within subjects design.

Why then, might we see differences between the two experiments? As touched on in the discussion section of the within subjects experiment (section 11.4) there are several methodological differences which might be causing variance between experiments.

The first is, of course, that the between subjects experiments compared two separate groups of readers. We did not anticipate that this would lead to systematic differences as the members of each group were selected at random from the same local population pool and were all subject to the same criteria (with the addition that those taking part in the chronologically second experiment (Load) had not participated in the NoLoad experiment; subject records were consulted to make sure that this criteria was fulfilled). There is no reason to expect that the participants in each group would differ from each other, but it is a factor we could not control for so may have contributed to our results in unexpected ways. Running a within subjects experiment to replicate the findings was the solution to this lack of control, but introduced other methodological differences of its own. Principal among them is the potential variance in overall task difficulty. The NoLoad group were required only to read sentences, there was no task switching involved in their experiment. The Load group were required to divide their attention between the digit and reading tasks, and switch between them within the each trial. The within subject participants were required to perform as the Load group did within trials, but also switch between NoLoad and load blocks, which may have provided an extra intensity. However, it meant that the within subjects participants were under load for approximately half the time that the Load group were (giving them breaks from the difficult task of dividing their attention), which is likely to have comparatively reduced the overall difficulty of the task. In general the results of the within subjects design indicate that readers were partially influenced by Load, but not as much as the Load group. This does tentatively fit with the idea that the within subjects experiment is providing a level of loading somewhere in between the two between subject groups.

As before, we have to consider whether participants might have been using strategies to combat the differing task demands of each experiment. Again, the data indicate that Load is influencing basic reading behaviour (more skipping, fewer mid-sentence regressions, more re-reading, longer early reading times) there is no obvious way to tell if this is due to a strategy or a true influence of Load. The fact that this behaviour replicates across both the load experiments in this section, and the one in the Syntax section do lend credence to the idea that it is a genuine effect, although it is not conclusive as it is possible that all our readers who were exposed to the digit load manipulation developed the same strategy. As before, however, measures were taken in the briefing of the experiments and the analysis of the comprehension questions / digit recall accuracy to limit strategic behaviour.

Similarly, we remain optimistic that our Load manipulation (digit span targeting the storage component of working memory) is having the desired effect in reducing the availability of working memory resources. We see Context, Ambiguity, and Load interacting in several places, as well as main effects of load in the within subjects experiment which are not attributable to between subject variance, all of which indicate that our manipulation is detrimental to interaction of the appropriate systems. It is possible, though, that our manipulation was not strong enough in the within subjects experiments, which is why we see limited support for the reduction of contextual influence on the processing of ambiguities. In order to test this idea we would need to include unambiguous controls.

Overall this series of experiments shows some support for the SLIR theory, as contextual influence on basic processing is not observably affected by an extrinsic memory load. However, we do see some interaction between Context, Ambiguity, and Load which neither the SLIR or SR theory would predict, although the effects are not consistent in locus or direction between experiments. As in the Syntax section, this hazy pattern of interaction results makes it hard to draw strong conclusions. It is interesting though, that in the Syntax section we saw more support for the SR theory as overall, normal processing appeared to be influenced detrimentally by Load. Procedurally speaking the experiments differed very little (identical presentation order, methodologies, experimental set up and location, etc.) but obviously there were differences within the materials used for each experiment. Both the type of ambiguity, and type of context used varied and it is initially tempting to focus on the ambiguity as the largest source of variance, as the context was conceptually the same (supportive or not). This is not a useful approach however as we are discussing variance in

the effect of Load on normal (non-ambiguous) processing between the two experimental series, and inherently that means discounting effects relating to type of ambiguity. The context then, differed in the sense that the lexical one was either neutral, which provided very little propositional content whatsoever, or supportive which created a much more propositionally full text regardless of the ambiguity. On the other hand, the syntactic one was specifically designed to provide (or not) a referential contrast set for the purposes of facilitating syntactic processing. It did have an influence above and beyond the ambiguity, in terms of discourse cohesion, but as was seen from the main effects, this influence was not as pronounced as that seen in the Lexical experiments. It is possible then, that our lexical experiments had contexts which had greater influence on normal reading behaviour, and so had more potential for disruption given the introduction of a memory load. This might account for why we found support for the SR theory in the syntax section, and the SLIR in the Lexical section.

It is possible then that experimental differences are accountable for our mixed findings, both between the Syntax and Lexical sections, and within the experiments of the Lexical section. Particularly, overall experimental difficulty might account for the failed replication of the Probability of Fixation interaction between the Lexical experiments.

Intrigued by the observation of this interaction, our final investigation into the use of Context while under Load was an examination into semantic prediction. This is reported in the next section.

## Section 4: Cognitive Load, Context, and Semantic Predictability

The history of prediction in language comprehension has been long and varied. In the early 1960s the prediction of upcoming words was assumed to occur as evidence was being observed that words presented under sub-optimal conditions that were contextually supported were identified faster and more accurately than those which were not supported contextually (e.g. Miller & Isard, 1963; Tulving & Gold, 1963). However, in the 1980s this concept was derided because, as the generative grammar of Chomsky (1957) and others implies, language is infinitely variable and so options open to the speaker are vast and task of predicting individual items seemingly impossible. This sentiment was summed up by Forster in 1981 and later reiterated by Jackendoff in 2002. These authors were considering the search space of a prediction in terms of word class, and syntactic constraints only, and were concerned that there were too many available options, and the cost of getting a prediction wrong and needing to reanalyse so large, that a predictive mechanism could not reduce overall processing costs, and was, therefore, simply not likely to exist. These concerns highlight two costs which are associated with a predictive mechanism: The first is the cost of making the prediction (with a minimally constrained search space leading to a huge number of potential options to choose from), and the cost of reanalysis when a prediction is proved wrong (much like a garden-path reanalysis). The benefits observed in earlier studies were considered to be side-effects of spreading activation (West & Stanovich, 1982) or ease of integration (Schwanenflugel & Shoben, 1985) rather than the result of a mechanism designed specifically to make predictions. Ease of integration (the reduction in the cost of integrating an item into the ongoing discourse if it is contextually supported) was for many years accepted as the mechanism which explained facilitation based on contextual information but was not empirically any more or less likely than a predictive mechanism. Rather, it was accepted because it explained the data, and fitted with the prevailing bottom-up approach of the time (mentioned above). A predictive mechanism on the other hand, also fitted the data but would have required a more top-down approach, allowing contextual information to constrain the search space beyond simply word-class or syntactic limitations and lexical activation to occur ahead of the predicted item.

Up until the mid-1990s the prevalence of eye-tracking as a methodology was limited, and the methodologies used were principally wholly or partially offline (such as grammatically judgments of self-paced reading). These methodologies provided data which was less temporally time-locked to stimuli than eye-tracking, which made the disentanglement of integration and prediction accounts very difficult. A more top-down approach to language processing became more popular in the mid-1990s, occurring around the same time as eye-tracking was becoming more accepted and showing contextual influence on reading and eye-movement behaviour. With this new approach, the idea of a predictive mechanism was reasserted and examinations attempting to separate integration and prediction became more common.

The literature pertaining to this shift in approach, and the work on prediction / integration is vast. We direct the reader to the work of Van Berkum, Brown, Zwitterlood, Kooijman, & Hagoort (2005) and Van Petten & Luka (2012) which provide comprehensive reviews. There are also several models attempting to specify precisely the mechanisms of context and sentence parsing but a detailed description of even one of these theories is beyond the scope of this thesis, and to discuss them all would be impractical in the extreme. Instead we refer the reader to Boston, Hale, Vasishth, & Kliegl, (2011) who provide a review as part of their introduction (see also Roland, Yun, Koenig, & Mauner (2012) for an briefer overview). As it is useful to have a framework in which to consider these concepts, however, we will briefly outline the idea behind one such model: A parallel, probability oriented, model known as Surprisal Theory (Hale, 2001; Levy, 2008).

In Surprisal Theory words are processed incrementally and the difficulty of processing a word is reduced as the probability of it occurring (given the preceding context) increases. Multiple possibilities are considered at once and the probability of all the possibilities currently being considered is always equal to 1 so that as one word increases in likelihood another must decrease. Therefore, when a reader encounters an unpredicted word its Surprisal value is low compared to the more predicted possibilities. Due to it requiring more 'effort' to select a word with a low Surprisal value (when other words with higher surprisal values are also being considered) readers find unpredictable words harder to process. In this way, the context automatically constrains the search space, and eases the processing of certain predictable words (via active prediction or ease of integration) and unpredicted words cause large processing difficulties.

Returning to our overview of Prediction / Integration: With the advent of eye-tracking, and particularly the visual-world paradigm, we began to see evidence of prediction occurring prior to the integration phase. Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy (1995) were one of the first to use the paradigm to show that the referents available to the participant could influence the way that they interpreted the syntax of a sentence (lending much support to the top-down rather than exclusively bottom-up perspective shift). Subsequently, Altmann & Kamide (1999) applied the methodology to investigate whether the semantic properties of referents would interact, online, with the semantic constraints of verbs. They presented a semi-realistic visual scene comprising of clip-art images representing items with various semantic properties (such as animacy). In the oft-given example, the referents included a ball, a cake, a boy, a toy car, and a toy train. The critical sentence 'the boy will eat the cake' was then presented audibly (with the control of 'the boy will move the cake'). The verb differs between the critical and control sentences only in the semantic properties which it dictates, i.e. 'eat' dictates an object which is edible, while 'move' is not constrained in such a way. Upon hearing the verb 'eat' participants made reliably more eye-movements to the only edible item in the scene (the cake), than any other, and, critically, reliably more eye-movements to the cake than when they heard 'move'. These eye-movements occurred before the word 'cake' was presented and so were termed anticipatory eye-movements. These anticipatory eye-movements provide strong evidence for prediction rather than integration as they indicate that participants were actively seeking out the object which fit with the semantic constraints of the verb, rather than simply finding it easier to integrate the eventual object once it had been revealed. Much research has followed this initial study, and Kamide provides an excellent review of this work (Kamide, 2008). As we are focusing on reading behaviour, however, we turn our attention to what is known about predictive eye-movements when reading text, rather than observing scenes.

It is worth noting that none of the studies mentioned here (regarding reading behaviour eye-movement control) are able to empirically make a distinction between integration and prediction effects.

One of the earliest studies to use eye-tracking to examine contextual effects on reading behaviour was Ehrlich & Rayner (1981). The authors presented two experiments in which they manipulated the preceding context and a critical noun. In the first experiment the

context was constructed to either strongly predict the noun, or be vague and have minimal impact on either the predictability or ease of integration of the noun. The noun varied in either being predicted by the context or not predicted by the context (and anomalous in the setting). It varied only by the alteration of one letter, however, so that the overall word shape could be manipulated as well; the altering letter could be at the start, middle, or end of the word and, depending where the change occurred, it would give a different visual profile to the word. Previous research (Ehrlich, 1981) showed that children were more likely to overlook changes if the context was highly constrained (possibly assuming they were spelling mistakes or not noticing them at all), but that when they did notice changes they were particularly sensitive to them at the beginning of the word and to those that changed the overall shape. In this experiment Ehrlich and Rayner showed a higher probability of skipping the highly predicted word (compared to that which wasn't supported by the context), and a dampened sensitivity to the alteration of one letter when the context was supportive. In the second experiment they again used the single letter change methodology but this time the word resulted in a word which was not anomalous with the context. They also removed the unsupportive context. This meant that they only varied the noun (by position letter difference) within the supportive context to be either predicted or not. In this simpler experiment they demonstrated that non-predicted non-anomalous words were fixated for less time than predicted anomalous words, and that orthographic features of the word were less heavily processed when the context was highly predictive.

Overall the relevant findings are such that readers are more likely to skip over words if they appear, on a surface level to fit with their contextually generated predictions, and if they do fixate on them, fixations are shorter for words that have been strongly predicted by the context. These findings are now well established, having been observed in a variety of studies since (e.g. McDonald & Shillcock, 2003; Traxler & Foss, 2000; Van Berkum et al., 2005). In addition, it is known that lexical frequency and predictability interact. Rayner, Ashby, Pollatsek, & Reichle (2004) conducted an experiment investigating the interaction of High and Low frequency words, which were either highly predictable or not predictable from the preceding context.

Before warming the milk, the babysitter took the infant's *bottle* out of the travel bag.

**High Frequency, Predicted**

To prevent a mess, the caregiver checked the baby's *bottle* before leaving.

**High Frequency, Unpredicted**

Before warming the milk, the babysitter took the infant's *diaper* out of the travel bag.

**Low Frequency, Predicted**

To prevent a mess, the caregiver checked the baby's *diaper* before leaving.

**Low Frequency, Unpredicted**

The authors argue that they found hints of an interaction between these two variables in early reading times such that low frequency words were more affected by predictability than high frequency words. However, this effect was not significant, and the numerical data indicates very small differences so it is to be considered with a great deal of caution. They did, however, find a significant interaction between prediction and frequency in the probability of fixating the critical word. This interaction patterned such that the High Frequency / Predictable condition showed a higher likelihood of being skipped than any of the other three conditions, all of which were practically identical (probability of fixation of 77% vs. 85%, 86%, 84%). See also Hand, Miellet, O'Donnell, & Sereno (2010) for a discussion of this paper and the lack of an interaction in fixation times it reports, and how the examination of the availability of para-foveal preview might indicate an interaction after all.

As mentioned above, these studies do not differentiate between prediction and integration accounts. Comparatively recently an attempt has been made to use ERP research to investigate this distinction. Below we will discuss two studies which have attempted to answer this question, before introducing our experimental work on the same topic.

In 2005 two studies were conducted investigating the prediction vs. integration distinction using Event Related Potentials (ERPs). ERPs are even more temporally accurate than eye-tracking and can give us an insight into the level of cognitive activity associated with each component of a sentence as it is read. The idea behind both these studies was that they,



much the same as the visual world paradigms, could examine processing of a prediction prior to the onset of the predicted word (when it becomes difficult to differentiate prediction and integration). Differently to the visual word studies though, these ERP experiments did not need to present a constrained set of potential referents, rather they could present a preceding context, much more equivalent to normal reading, that provided the constraint required. Other studies by Wicha and colleagues (Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2004) are not discussed here but follow the same general idea.

The first is Van Berkum et al. (2005) who employed the gender markings of Dutch adjectives to disconfirm sentential predictions prior to the actual noun being encountered. In auditory presentation, participants received a preceding context sentence, followed by either a predicted or unpredicted target (as below).

**Context Sentence:** The burglar had no trouble locating the secret family safe.

**Predicted Target:** Of course, it was situated behind a big but unobtrusive painting.

**Unpredicted Target:** Of course, it was situated behind a big but unobtrusive bookcase.

In Dutch, the adjective 'big' is gender-marked as either common or neuter depending on the noun. In this example the noun 'painting' is gender neuter, while the noun 'bookcase' is gender common. Therefore, on encountering the gender-marked adjective 'big' in the unpredicted 'bookcase' condition, participants know that their prediction for the safe to be hidden behind a painting is incorrect. This was predicted to show up in the ERP record. All items were subjected to a pre-test in which Cloze values (the likelihood of a person finishing off the sentence with the particular word) were acquired. This allowed Van Berkum et al. to identify which items would allow for the greatest effect when the prediction was violated. They found that a small positivity was found 329 milliseconds (on average) after the onset of the adjective, and a later N400 signature 350-400 milliseconds after the noun onset. They interpreted these results as evidence that participants listening to sentences are sensitive to morpho-syntactic information which counters their predictions prior to the predicted word being encountered, as well the anticipated violation

of their prediction / ease of integration effect on the noun itself. In the same series, Van Berkum et al. conducted a self-paced reading experiment which showed comparatively delayed, but still present, increase in reading times prior to the noun onset. In this case the slowdown did not manifest until the second inconsistent adjective although Van Berkum remain agnostic as to whether that represents spillover from the first inconsistent adjective, an accumulation of inconsistent evidence finally tipping the balance, or something specific about the second adjective as compared to the first. Nevertheless, this self-paced reading study appears to indicate evidence to support their main conclusions.

The second study was conducted by DeLong, Urbach, & Kutas in 2005. DeLong et al.'s idea was similar to that of Van Berkum et al. but instead of Dutch adjectival gender markings they made use of the phonological regularity in English which dictates that indefinite articles preceding a word adapt their form dependant on whether that word begins with a vowel or consonant sound (e.g. an elephant, an apple, an ornament, a pencil, a dog, a telephone). Semantically speaking, the articles 'a' and 'an' are identical, but they vary in phonology. By contrasting highly constraining sentences with an article which does or does not fit with the word predicted by the context, DeLong et al. hoped to observe ERP fluctuations prior to the onset of the critical word. The sentences took the form given in the example below.

The day was breezy so the boy went outside to fly *a kite* in the park.

**Predicted 'a'**

The day was breezy so the boy went outside to fly *an airplane* in the park.

**Unpredicted 'an'**

Although the idea of flight was as old as the hills, the Wright Brothers were the first people to build *an airplane* that actually flew.

**Predicted 'an'**

Although the idea of flight was as old as the hills, the Wright Brothers were the first people to build *a kite* that actually flew.

**Unpredicted 'a'**

Like Van Berkum et al., DeLong et al. conducted a pre-test to establish Cloze values for each of their items. In contrast to Van Berkum et al., they presented the sentences visually rather than using an auditory presentation. DeLong et al. measured ERPs on the noun (kite/airplane) and the article (a / an) with the idea that under a prediction account they

would observe an effect on the article (as the prediction was violated) but under an integration account, only the noun would generate an ERP response as no prediction is being violated, it is the difficulty of integrating the noun which is being modulated by a constraining context. They found that their materials elicited the expected correlation between N400 amplitude and the offline cloze probability of the noun (i.e. a smaller N400 when the noun is predicted compared to a large N400 when it is unpredicted). In terms of the articles they found a smaller but similar (and significant) effect. If the phonological form of the article did not agree with the predicted noun then a greater N400 signature was observed than if it did. The authors argue that this provides further evidence of online prediction and detracts from integration accounts. We note, however, that methodological reasons might have partially or wholly caused the N400 effect observed on the article. Our principal concerns are a). the timing of this effect: DeLong et al. presented the article for an extended duration (200ms with a 500ms onset asynchrony), which far exceeds the duration of a normal fixation on a function word in reading (the average reading time of a word in silent reading is 225ms but this varies with length such that shorter words have receive shorter fixation times (Rayner, 1998)), and b). the constrained nature of the reading occurring in the ERP study (i.e. no facility for preview or skipping). Although DeLong et al. argue that there is no reason to assume variation in the neural mechanisms involved in reading, we remain sceptical that such constrained circumstances accurately reflect normal reading behaviour.

It is for this reason that, in the experiment we present below, we attempt to replicate the findings of DeLong et al. using an eye-tracking methodology. Additionally, in keeping with our general theme we will also be putting participants under cognitive load to observe how a reduction in resources influences the systems responsible for semantic prediction.

There are, to our knowledge, no studies directly investigating how resource reduction impacts predictive mechanisms, but there is evidence from the computational linguistics / parsing literature that limiting the resources available to a parallel processor so that fewer parallel representations can be maintained at once appears to reduce the functionality of the parser (e.g. Boston et al., 2011; Levy, 2008; Lewis, Vasishth, & Van Dyke, 2006; Lewis & Vasishth, 2005). Regardless, since these mechanisms appear to be inherent in the processing of language, and are essentially correlates of the unambiguous conditions (or

'normal' processing) of our previous experiments, the overarching SL and SLIR theories remain applicable.

The SR theory posits a single resource pool from which all language oriented cognitive processes draw. As such it predicts that a reduction in resources will negatively affect all processes, and particularly the interactions of process (in this case, context processing and prediction / integration). Therefore under load we should see no difference between contextually highly predictable and unpredictable words, as the use of contextual information in generating Surprisal values will be reduced or eliminated. When not Loaded we should see the established difficulty for processing unpredicted words. The SLIR theory on the other hand, maintains that at least two resource pools exist, one for interpretive processes (normal processing under which parsing / prediction / integration fall), and one for post-interpretive processes (conscious manipulation of verbal information like ambiguity resolution or the remembrance of lists). The argument is that the suppression of resources in one pool will not affect processing which draw from the other. Critically, our Load manipulation (assuming this perspective) taps post-interpretive resources and not interpretive resources. Given that parsing / prediction / integration falls within the remit of interpretive resources, our load manipulation should have no effect and we should see an unchanged pattern.

## **Chapter 13: Load, Context, and Semantic Predictability – Within Subjects (Experiment 6)**

### **13.1 Introduction**

Our observations in previous experiments (Section 3) of how cognitive load might influence the use of context in the pre-processing of critical words have led us to consider the interaction between load and context in guiding fixations in normal parsing. This chapter reports a single within subjects experiment investigating the influence of cognitive load on the use of context to facilitate processing of predictable words. In particular, we will be attempting to replicate DeLong et al. (2005) who showed evidence of a predictive mechanism by manipulating the preceding context and the phonological form of the indefinite article in English (a / an), and subsequently observing ERP signatures prior to the onset of the contextually predicted noun (which the article preceded) that indicated a violation of expectation. Using a subset of DeLong et al.'s cloze tested materials, which were altered slightly to better fit our population, we aim to replicate this finding using eye-tracking as a simpler methodology which is closer to normal reading than the single word, slowed presentation of ERPs. We also add two further factors: the first is lexical frequency as there is much discussion in the literature regarding additive or interactive effects of frequency and predictability, and the second is our Cognitive Load manipulation which we have used in previous experiments. There are therefore three main threads to our examination. There is no direct mapping of ERP measures to eye-tracking measures but we predict a manifestation of prediction and load effects / interactions caused by 1). the phonological variation of the a / an article, 2). the interaction of frequency and predictability on or around the critical word, and 3). the mitigation of both 1). and 2). by the application of cognitive load. We discuss each of these in turn below.

First, the critical prediction vs. integration manipulation of the a / an article. DeLong et al. presented a preceding context which supported either a head noun which started with a consonant or a vowel. In English there is a regularity that states that the phonological form of an indefinite article varies based on the onset sound of the subsequent noun: An precedes vowel sounds (e.g. an airplane) while A precedes consonant sounds (e.g. a kite). They found a small but significant N400 signature on the article which was modulated by

predictability such that when a stronger prediction was violated, a greater negativity was observed. The presentation time for the article was 200ms with a 500ms onset asynchrony. This allowed for the negativity to be observed 400ms after presentation. In eyetracking the presentation of the sentence is not incrementally compartmentalised, so the reader can spend as much or as little time on the article as they require (or can skip it entirely). Following the findings of DeLong et al. we anticipate three possible outcomes. Firstly, longer fixation times on the article if it violates the prediction generated by the context. Secondly, a probability of fixation effect on the violating article such that it will either be skipped more in an attempt to confirm the violation (attraction effect) or skipped less if it represents enough information to fully disconfirm the violation in itself. This would be considered evidence of prediction only if the information about the noun was not available para-foveally. In either case we might expect to see a higher probability of fixating the noun if the article indicates a violation. Thirdly, there may not be enough time in the average fixation duration of an article for the parser to realise the violation (although the established concept of an incremental parser would indicate that the next fixation should not occur until all processing on the article is complete), and so reading behaviour may be unaffected.

The interaction of frequency and predictability on or around the critical word is not overly theoretically driving for the research we present here. We have, in the past experiments, included frequency only as a way to account for variance in our data which we do not want to miss-credit to our main manipulations. In the realm of predictive parsing the predictability and frequency discussion is ongoing and so we will mention any pertinent findings, even though our experiment was not designed to investigate this phenomenon directly. Rayner et al. (2004) showed no interaction between the effects of predictability and frequency on fixation durations in early reading measures, but did find that the two variables interacted in the likelihood of fixating the critical word. They found that highly predictable words which were also of high frequency were skipped more often than any other of the possible conditions (which were all the same). Hand et al. (2010) replicated this probability of fixation interaction, and also found evidence to indicate that the amount of para-foveal preview (as indexed by launch site from the previous word to the critical word) influenced the duration of fixations on the critical word in a conditionally interactive pattern. However, as we are not performing a launch site analysis it is unlikely that we will observe any interactions in early fixation durations on the critical word. We do, however,

anticipate the predictability by frequency interaction in the probability of fixating the critical word, which both the above mentioned studies have shown.

As mentioned at the end of the introduction to this section, the SR and SLIR theories are applicable to this experiment. Our Cognitive Load manipulation reduces the resources which the SLIR theory considers to be part of the post-interpretive resource pool, but the making of predictions / ease of integration issues we are investigating are the purview of the interpretive resource pool. The SLIR theory therefore predicts no interaction between load and predictability. The SR on the other hand (which posits only one resource pool) predicts an interaction between predictability and load such that load reduces the positive effect of a preceding context making a word more predictable or, to look at it another way, reduces the negative effect of a violation because the context was not used to provide evidence in favour of any particular word. Therefore, a reduction in the effects of the a / an manipulation (discussed above) would lend support to the SR theory, while no difference between the Load and NoLoad conditions would advocate the SLIR theory. With regards to a predictability by frequency interaction, the same applies. Both frequency and predictability are processes which the SLIR theory would assume draw from the interpretive resource pool and so should remain unaffected by our load manipulation, whereas the SR theory would predict a general degradation of interactivity between processes, and so would predict an interaction with load.

## 13.2 Methods

### *Participants*

36 participants meeting the following experimental criteria took part in the study. All were native speakers of British English, over 18 years of age had no known reading disorders (e.g. dyslexia), and had normal or corrected to normal vision.

Total Exclusions based on poor comprehension, tracking, or digit recall = 1

Comprehension Question Range (of included participants) = 83% - 97%, Mean = 92%.

Digit Span Range (of included participants) = 5 – 9, Mean = 6.

### *Stimuli*

40 of the stimuli from the original DeLong et al. (2005) paper were selected based on their appropriateness for our population. Specifically, items including Americanisms, references to American cultural items, and words with overtly American interpretations (e.g. elevator) were excluded (see Appendix 3 for a full list of items). As with the DeLong paper we utilised the phonological difference (but semantic same-ness) of the 'a' / 'an' Articles to examine predictability versus integration. This required us to have four variations of each item: an 'a' predicted version, an 'an' unpredicted version, an 'a' unpredicted, and an 'an' predicted version. We therefore had 20 items with four versions each. The context sentences and critical nouns were different between two of those four versions, however, so we were able to present two versions to each participant without either component being repeated. A / An's were distributed equally overall (so that half the items contained 'a', and half contained 'an'), and each fitted with a contextual predicted noun in half of the items, and an contextually unpredicted noun in the other half. A set of example stimuli is given in 13.2.a - 13.2.d below.

This experiment was run as 'fillers' within the Lexical Within Subjects Load experiment reported in the previous experimental section (Chapter 11), therefore the same mini-blocking system and pseudo randomisation approach was adopted. Blocking was four items deep and alternated between NoLoad and Load by block. Within a block, all items were randomly selected from their lists. The first item was filler, the second was an experimental item from the lexical experiment, the third was an experimental item from this experiment, and the fourth was another filler. This meant that the state of Loading did not change on an experimental item, and that the experimental items from each experiment were always a minimum of four trials apart. See Chapter 11: (11.2.e) for an example of the mini-blocking system.

In the DeLong et al. study, lexical frequency of the head nouns was not controlled across conditions (e.g. Carrot has a different lexical frequency to Apple). Nevertheless, the counterbalancing of the experiment dictates that each head noun will appear in each condition (i.e. contextually predicted or not, while under load or not) and so comparison across conditions based on frequency is still possible.



13.2.a.

For the snowman's eyes the children used two pieces of coal, and for its nose they used a carrot from the fridge.

(‘a’ Predicted)

13.2.b.

For the snowman's eyes the children used two pieces of coal, and for its nose they used an apple from the fridge.

(‘an’ Unpredicted)

13.2.c.

The old wives tale says that if you want to keep the doctor away then you should eat an apple a day.

(‘an’ Predicted)

13.2.d.

The old wives tale says that if you want to keep the doctor away then you should eat a carrot a day.

(‘a’ Unpredicted)

### *Procedure*

The procedure of this experiment was identical to that of the Within Subjects Load experiment reported in Chapter 11. The same digit span pre-test and pre-rest procedures were used to tailor the level of difficulty to each participant. It was explained that the experiment would only have a Digit component for half of the items. For the blocks without Loading, the presentation and procedure was identical to that of experiment 3, for the blocks with Loading, the presentation and procedure was identical to that of experiment 4.

### *Data Analysis*

Three factors were created: Predictability (whether the noun was predictable from the context or not), Load (whether the participant was under the memory load or not), and we also used the British National Corpus' frequency values to establish Frequency (a categorical predictor categorising the lexical frequency of the main noun as high or Low frequency – see Chapter 8: (8.2) for details). We therefore have a 2x2x2 between Predictability, Load, and Frequency.

The experimental items were broken down into six regions. The first region comprised the context sentence (Context region). This was always contained on one line, with a line break separating it and the following target regions. The target text was broken down into five regions. The first included all words up to the word preceding the critical noun (Initial region). The second was the article preceding the noun (Article region). The third was the main noun (Noun region). The fourth region was the word following the critical region (Spillover region). The fifth was the rest of the sentence (Wrapup region). An example of the region breakdown is given in 13.2.e:

#### 13.2.e. Region Breakdown

```
/ For the snowman's eyes the children used two pieces of coal, and for its./  
/                                     1                                     /  
/ nose they used/ a / carrot/ from/ the fridge./  
/      2      / 3/ 4 / 5 / 6      /
```

1 = Context, 2 = Initial, 3 = Article, 4 = Noun, 5 = Spillover, 6 = Wrapup

An additional region was constructed by combining 3 and 4 (the Article and the main noun). This region will be referred to as the combined Article+Noun region.

### 13.3 Results

/ nose they used/ a / carrot/ from/ the fridge./  
/        2        / 3 / 4 / 5 / 6        /

1 = Context, 2 = Initial, 3 = Article, 4 = Critical Noun, 5 = Spillover, 6 = Wrapup

#### Overview of Critical Findings:

There are two main findings of this experiment:

A marginal ( $p = 0.059^m$ ) Predictability by Load interaction in the probability of fixating the noun region. For readers under Load, no difference is observed across predictability, while readers who are not under Load are more likely to fixate the noun if their prediction is violated. The predicted conditions are the same across Load which implies that Load does not inhibit normal use of context to create predictions. The difference is only observed in the unpredicted conditions which implies that those under load are less able to adapt to having their predictions violated. See Figure 13.1 for details.

A marginal ( $p = 0.068^m$ ) Frequency by Predictability interaction in the First Pass of the Article+Noun region such that Low Frequency words are unaffected by Predictability, but reading times on High Frequency words are reduced when the noun is Predicted from the context (compared to when it is not). This interaction is not observed by Rayner et al. (2004) or Hand et al. (2010) although it patterns the same way as the probability of fixating the noun Frequency by Predictability interaction that they report. See Figure 13.2 for details.

#### Predictability by Load interaction:

Probability of Fixating the Critical Region:

We observe a marginal interaction between Predictability and Cognitive Load in the probability of fixating the Noun region ( $p = 0.059^m$ ) (Figure 13.1). There is no statistical difference between Predictability conditions within Load conditions (NoLoad / Unpredicted = 70% vs. NoLoad / Predicted = 67 %,  $p > 0.1$ ) (Load / Unpredicted = 63% vs. Load /

Predicted = 67%,  $p > 0.1$ ), or across Load for the Predicted condition (NoLoad = 67% vs. Load = 67%,  $p > 0.1$ ) but there is a statistical difference across Load for the Unpredicted condition (NoLoad = 70% vs. Load = 63%,  $p = 0.007^{**}$ ). Initially it appears that readers under Load are not able to make use of predictive information in the same way as those not under Load. This is supported by the fact that there is no statistical (and a very small numerical) difference between the Predicted and Unpredicted conditions for the Loaded readers. Critically, there is a larger numerical difference in the likelihood of fixation when the word is Unpredicted (compared to Predicted) for those not under Load, which fits with the anticipated pattern. Although larger numerically, this effect is also not significant. This interpretation would support the SR theory as it shows the expected pattern of predictability occurring for those not under Load, but no effect of prediction for those under Load. The interaction is marginal, however, and while the data trends suggest this interpretation is valid, the significance of the prediction effect in the NoLoad condition would be critical to being able to assert this perspective with confidence. It is not significant (or even marginal) so we must consider other alternatives.

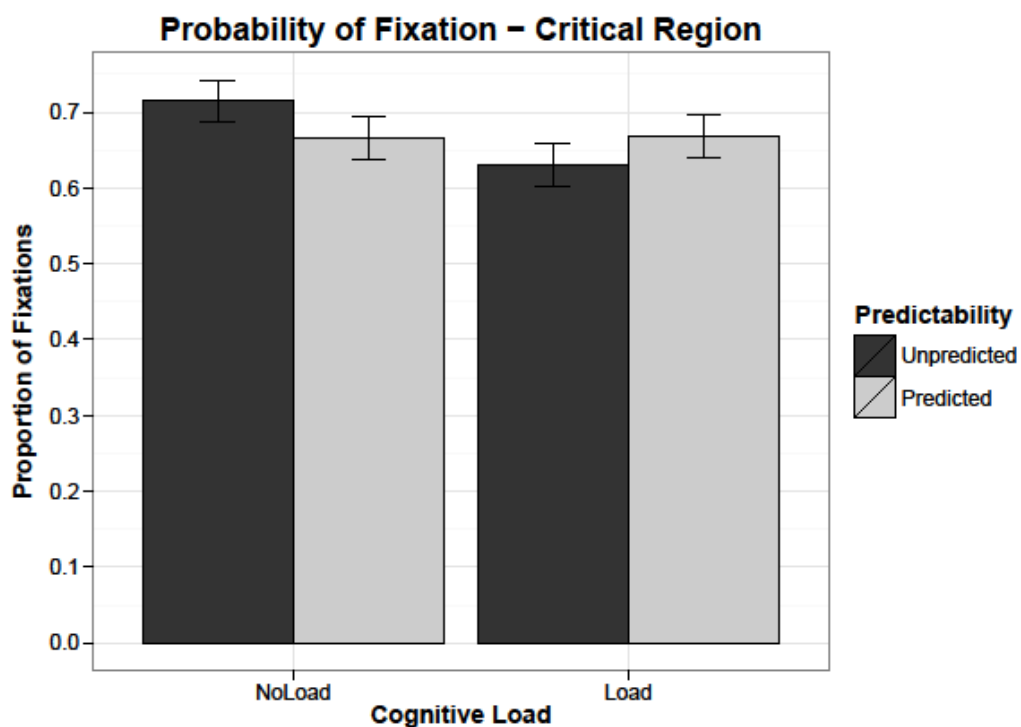


Figure 13.1: Experiment 6 - Proportion of Fixations for Probability of Fixation on the Critical Region. Error Bars display Standard Error. Marginal Predictability by Load Interaction.

A second interpretation is that readers under load are still able to use predictive information (supported by the two Predicted conditions being almost identical across Load) but they are not as readily able to notice their predictions being proved wrong (either by the presence of an incongruous article or para-foveal word identification) or adapt as quickly to it happening: When the upcoming word is not the one they predicted, readers who are not under Load are more likely to fixate it than those under Load. This may be because they are more sensitive to lexico-syntactic information (a / an noun agreement) or to information in the para-fovea and so can identify when their prediction is about to be proved incorrect, or it may be that they have more available resources with which to adapt their eye-movement plan on the fly. Either way, this second interpretation appears to fit better with the data and suggests that Predictive mechanisms still function normally under Load, but the processing following a violation of a prediction does not. This provides support for the SLIR theory as normal contextual influence is maintained, but in cases of violation, when more conscious attention is oriented onto solving the anomaly (just as in ambiguity processing) the limitation of post-interpretive resources is evident. The interaction is Marginal, however, so as always should be considered with caution.

#### **Predictability by Frequency Interaction:**

First Pass of the Article+Noun Region:

In First pass we observe a marginal Predictability by Frequency interaction which we did not anticipate. Rayner et al. (2004) showed no evidence of an interaction between these two factors on the critical noun (finding instead two additive main effects), and Hand et al. (2010) showed some mixed evidence of an interaction but only after including launch site (a measure of the amount of para-foveal preview) as a factor.

Here we observe a marginal interaction ( $p = 0.068^m$ ) on the combined Article+Noun region (Figure 13.2). There are no significant differences between three of the conditions (LowFreq: UnPred = 257ms, Pred = 259ms, HighFreq: Unpred = 261ms) but the Predicted / High Frequency words receive significantly shorter fixation time (240ms) than their Low Frequency counterparts ( $p = 0.030^*$ ).

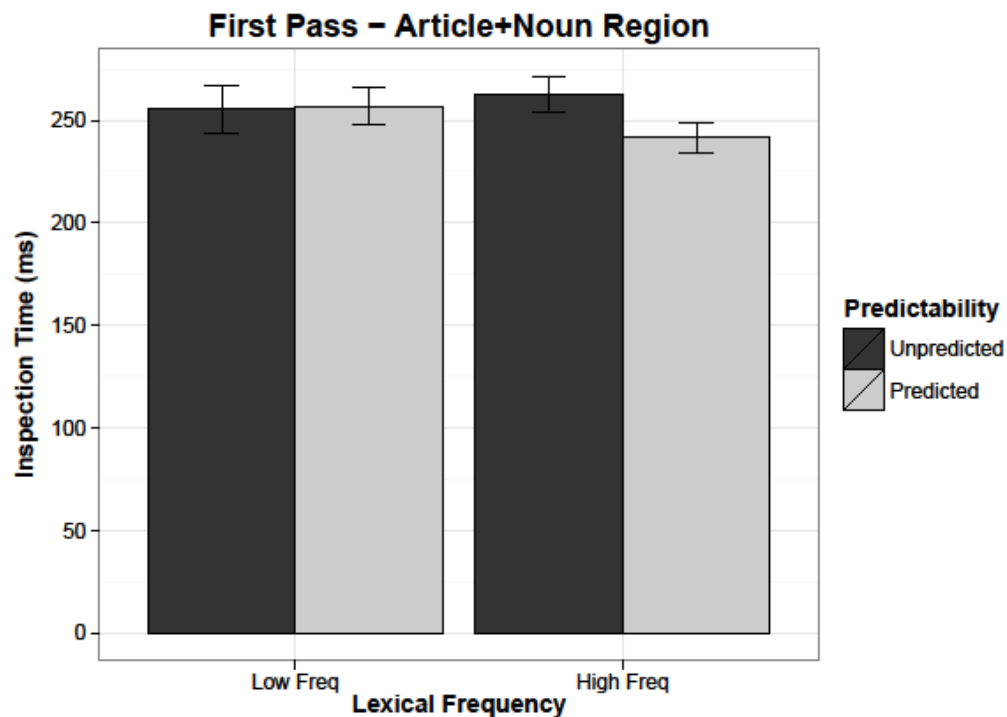


Figure 13.2: Experiment 6 – Mean inspection time for First Pass on the Article+Noun Region. Error Bars display Standard Error. Predictability by Frequency Interaction.

This replicates, in an early reading measure, the pattern that Rayner et al. (2004) and Hand et al. (2010) found for the probability of fixating the critical noun. This interactive pattern is replicated in several other measures throughout our analysis, including significantly in the Regression Path of both the Article+Noun and Spillover regions, the Total Time spent in the Wrapup region, the Regressions Out of the Noun region, and the Regressions In to both the Initial and Noun regions (all  $p$ 's < 0.05 \*). This pattern implies that there is a reduction in difficulty (shorter reading times, and fewer regressions out and in) for the high frequency predictable words, compared to all other conditions. This is perhaps unsurprising, but it is notable that while we did find this interaction in multiple places (and in doing so have extended the probability of fixating the noun effect which both Rayner et al. and Hand et al. have shown), we did not find it where we expected to: Our analysis was conspicuously missing a prediction by frequency interaction in the probability of fixating the noun. Possible reasons why we found this interaction where Rayner et al. and Hand et al. did not (without launch site analysis), but did not find it where they did, will be discussed in the upcoming discussion (section 13.4).

### **Other Interactions:**

Below we will report any other relevant interactions. As in previous sections there were many interactions which involved the control factor Frequency. Table 13.1 lists interactions involving frequency which will not be discussed. It is worth noting, however, that we will discuss selected interactions involving Frequency, and that unless specifically stated, all those listed in Table 13.1 pattern the same way as those discussed.

To anticipate our other results we find:

Load by Frequency Interactions:

#### *Regression Path – Noun Region*

Regression Path times for High frequency words are unaffected by Loading. Regression path times for Low frequency words on the other hand are significantly inflated when readers are placed under Load (see Figure 13.3). All observations of this interaction pattern the same except First Fixation in the Spillover Region.

#### *First Fixation – Spillover Region*

This interaction patterns differently to all others of its type. This is the very first fixation made after exiting the noun region rightwards and it shows inflated fixation times for the NoLoad / Unpredicted condition, with all other conditions being statistically equal (see Figure 13.4).

Table 13.1: Experiment 6 - Interactions involving Frequency which are not reported in the Results section.

Factors	Measure	Region
Load + Freq	Regression Path	Article+Noun
	First Fixation	Spillover
	Regressions Out	Critical
Predictability + Freq	Total Time	Wrapup
	Regression Path	Article+Noun, Spillover
	Regressions Out	Noun
	Regressions In	Initial, Noun

#### Load by Frequency Interactions:

Regression Path for the Critical Noun Region:

For the Noun region we see a Load by Frequency interaction ( $p = 0.009^{**}$ ) in which High Frequency words are unaffected by Load (NoLoad = 384ms vs. Load = 379ms,  $p > 0.1$ ) (see Figure 13.3). Within Load conditions there also no statistically significant differences although we do see opposing directions for the numerical trends (NoLoad / Low Freq = 304ms vs. Load / Low Freq = 384ms,  $p > 0.1$ ) (Load / Low Freq = 459ms vs. Load / High Freq = 379ms,  $p > 0.1$ ). There is a significant difference across Load conditions, however, for Low Frequency words (NoLoad = 304ms vs. Load = 459ms,  $p = 0.0001^{***}$ ). It seems, then, that the processing of High frequency words is unaffected by whether participants are under Load or not, while Low frequency words present a much a greater challenge to Loaded readers than to non-resource limited readers. This pattern is replicated across all other instances of this interaction except one observed in the First Fixation of the Spillover region, which is described below.



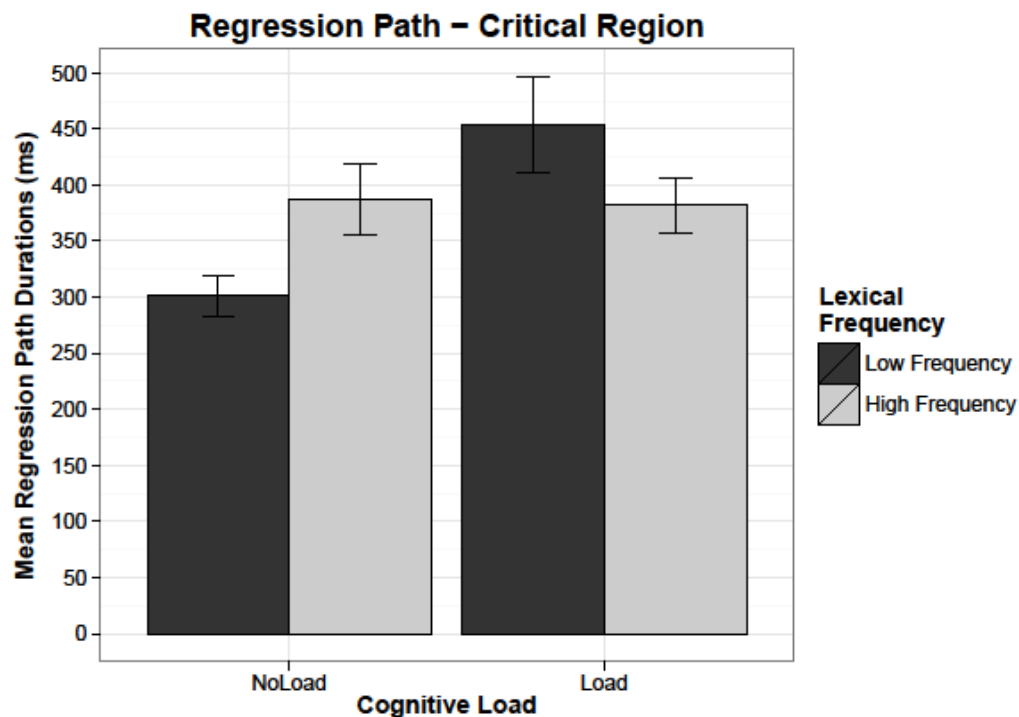


Figure 13.3: Experiment 6 – Mean Regression Path Duration for Regression Path for the Critical Region. Error Bars display Standard Error. Load by Frequency Interaction.

#### First Fixation on the Spillover Region:

In the Spillover region we see a different pattern for the interaction reported above (Load by Frequency). Here (Figure 13.4,  $p = 0.007^{**}$ ) we see no statistical differences within the Load condition (Low Freq = 236 vs. High Freq = 245ms,  $p > 0.1$ ) and no statistical differences across Load for Low Frequency words (NoLoad = 241ms vs. Load = 236ms,  $p > 0.1$ ). We see a marginal difference across Load for High Frequency words (NoLoad = 223ms vs. Load = 245ms,  $p > 0.1$ ) such fixation durations are shorter in the NoLoad condition. Additionally, we see a statistical difference across Frequency in the NoLoad condition (Low Freq = 241ms vs. 222ms,  $p = 0.03^{*}$ ). It seems that there is a suppression of first fixation times in the NoLoad / High Frequency condition, which is driving the significant difference between Low and High frequency words in the NoLoad condition. Thus participants are spending less time in the fixation immediately after the critical region if the previous word was of high frequency (compared to Low frequency, but only if they are not under Load. It is not obvious why this might be, but one simple explanation is that this is the 'easiest' condition,

and a). readers may simply find it easier than any other, or b). may have had sufficient time on the previous region to engage in para-foveal processing and so need less time here.

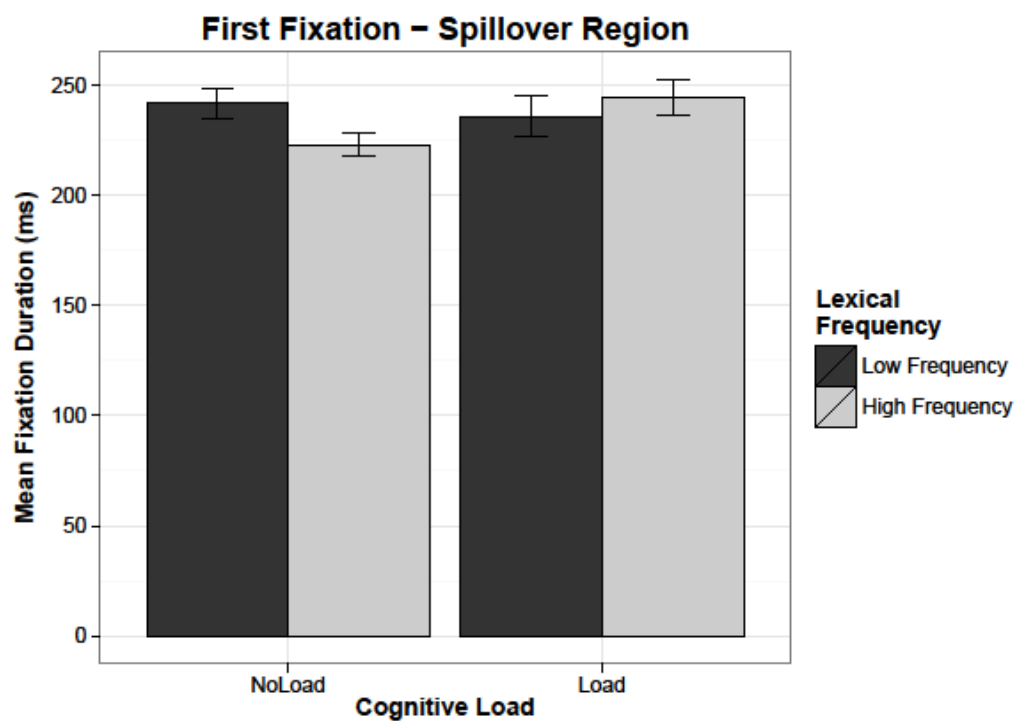


Figure 13.4: Experiment 6 – Mean Fixation Durations for First Fixation of the Critical Region. Error Bars display Standard Error. Load by Frequency Interaction.

### **Summary of Main Effects:**

The main effects are reported in the tables below. Table 13.2 reports Context effects, while Table 13.3 reports Ambiguity effects.

#### **Predictability:**

The picture presented by Table 13.2 is very clear. In all cases reading times, fixation durations, probability of fixation, and proportions of regressions are reduced if the critical noun is predicted given the preceding context. This effect is observable on the Noun in first pass and then subsequently in all other regions during re-readings (SP). Of interest is that we do not see this effect on the Article in first pass, despite reading times being reasonably long for an Article. It was fixated much less than the Noun however, so it is possible that we do not have sufficient observations for any differences to be observed.

#### **Load:**

Load is much more varied than the homogenous main effects of Predictability (see Table 13.3). Effects on the Noun are consistent such that while under Load participants make longer fixations and have longer reading times and regression path times than when not under Load. Conversely, those under load appear to spend less time in the spillover region, but make more regressions out of it, and spend longer within those regressions. One other interesting observation is that those under Load are less likely to fixate the initial region, and will spend less time there (FP) if they do. This indicates a rush to get to the more critical parts of the sentence, which fits with the hypothesis presented earlier of faster and shallower processing by those under Load.

Table 13.2: Experiment 6 - Main Effects of Predictability. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if the Noun is Unpredicted or Predicted by the preceding context. Effects are also marked if Marginal.

<i>Predictability</i>	<b>Levels</b>	<b>Initial</b>		<b>Article</b>		<b>Noun</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Article+Noun</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	Unpredicted	0.39	488.2	-0.14	223.2	-1.98	231.4	0.56	335.7	-0.21	400.1	-1.34	259.6
	Predicted		522.0		220.4	*	217.4		344.8		396.3		249.6
First Fixation	Unpredicted	0.87	202.9	-0.22	222.5	-1.17	208.9	-1.04	237.1	-0.36	254.1	-0.56	212.7
	Predicted		206.7		218.1		204.4		235.0		246.8		210.8
Regression Path	Unpredicted	-0.32	571.7	0.40	307.1	-2.29	398.7	-2.01	826.9	-3.45	1574	-1.36	411.6
	Predicted		586.7		341.1	*	364.5	*	765.8	***	1225		406.0
Second Pass	Unpredicted	-3.22	234.9	-2.52	52.44	-4.43	132.8	-3.19	185.2	-2.38	114.69	-2.52	178.3
	Predicted	**	177.0	**	38.84	***	76.32	***	127.98	**	73.70	**	110.8
Total Time	Unpredicted	-3.67	669.0	-0.27	257.5	-5.86	360.9	-3.04	526.1	-2.30	563.6	-6.09	429.6
	Predicted	***	638.2		250.6	***	287.0	**	469.0	*	503.0	***	346.4
Probability of Fixation	Unpredicted	-2.13	.73	1.13	.22	-0.38	.68	-3.66	.82	-0.54	.70	-0.26	.82
	Predicted	*	.69		.24		.67	***	.73		.69		.81
Regressions Out	Unpredicted	-1.12	.06	-1.01	.32	-1.59	.33	-1.94	.40	-2.04	.76	-1.79	.29
	Predicted		.05		.29		.30	m	.36	*	.70		.27
Regressions In	Unpredicted	-4.75	.49	-2.89	.20	-4.80	.33	-2.34	.28	/	/	-4.94	.39
	Predicted	***	.38	**	.15	***	.23	*	.23		/	***	.28

Table 13.3: Experiment 6 - Main Effects of Load. T Values / Z Scores and Means for all Regions. Colours display direction of effect: Shorter Reading Times & Lower Proportions of Fixations if participants are under **Load** or **NoLoad**. Effects are also marked if **Marginal**.

<b>Load</b>	<b>Levels</b>	<b>Initial</b>		<b>Article</b>		<b>Noun</b>		<b>Spillover</b>		<b>Wrapup</b>		<b>Article+Noun</b>	
		T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means	T / Z	Means
First Pass	NoLoad	-3.27	507.0	-1.38	226.8	3.87	213.1	-2.60	347.6	1.17	372.3	1.32	248.0
	Load	**	503.2		216.8	***	235.6	**	332.9		424.1		261.2
First Fixation	NoLoad	0.87	201.3	-1.45	226.1	4.40	196.4	0.50	232.0	1.64	240.8	2.08	206.9
	Load		208.3		214.5	***	216.9		240.1		260.1	*	216.6
Regression Path	NoLoad	-1.51	557.9	-1.33	339.1	3.25	343.5	2.43	692.4	0.69	1424	1.59	383.4
	Load		600.5		309.1	**	419.7	*	900.3		1375		434.1
Second Pass	NoLoad	1.36	188.4	-0.33	46.63	0.98	97.00	-0.16	158.2	-0.15	97.41	-0.33	137.8
	Load		223.5		44.66		112.1		155.0		90.98		151.4
Total Time	NoLoad	-1.33	636.8	-1.02	260.3	2.74	305.0	-1.99	503.4	0.80	505.6	1.59	373.9
	Load		670.4		247.7	**	342.9	*	491.7		561.0		402.1
Probability of Fixation	NoLoad	-4.74	.74	-1.49	.25	-1.66	.69	-1.51	.79	-2.57	.72	-2.63	.84
	Load	***	.67		.21		.65		.76	*	.67		.79
Regressions Out	NoLoad	1.82	.04	-1.00	.33	0.36	.30	3.37	.34	-1.37	.75	-0.10	.28
	Load	<sup>m</sup>	.07		.27		.33	***	.42		.71		.28
Regressions In	NoLoad	-0.74	.44	-0.12	.18	-0.32	.28	-2.33	.28	/	/	0.16	.34
	Load		.44		.18		.27	*	.23		/		.34

## 13.4 Discussion

This experiment was conducted with the view of replicating the main findings of DeLong et al. (2005) using an eye-tracking methodology rather than ERP, and to extend those findings to see how predictive mechanisms might be effected by the application of Cognitive Load. Ultimately, we hoped that in doing so we could gain further evidence to support one of the two principal theories of cognitive resource systems: the Single Resource theory or the Separate Language Interpretation Resource theory.

Firstly, we will discuss our findings in a non-resource restricted environment with reference to the predictions made based on the ERP findings of DeLong et al. (2005) and the eye-tracking findings of Rayner et al. (2004) and Hand et al. (2010).

We found longer early reading times on the critical noun if it was unpredicted by the context, which was predicted by both the ERP and eye-tracking data. This finding is not surprising, and fits well with what is already known about violating predictions. Another predicted finding which we did not observe, however, was the skipping of the article or the noun. We anticipated some kind of skipping effect on the article (whether it were a greater likelihood of fixation due to the violation of a prediction or a higher chance of skipping due to an attraction effect to the noun to confirm the violation). We saw no evidence of a main effect of Predictability on the article, which implies that readers are unable to make use of the phonological clue provided by the a / an regularity in time to alter their saccadic plans. DeLong et al. argue that their ERP results, with such an inherently slow presentation rate, should map onto other presentation rates as the neural mechanisms underlying the findings should not differ by presentation rate. They cite Kutas (1987) and Gunter, Jackson, & Mulder (1992) as examples of research into presentation rate which support their claim. However, we believe there may be other factors at play here which sidestep this claim. We see two possibilities as to why we do not observe the expected effect. Firstly, while it may be the case that the underlying neural mechanism providing their ERP results is the same as that used in more normal reading, the application of time pressures from the need to generate and execute a saccadic plan may limit the amount of inference that can take place in terms of information uptake and integration; is there sufficient time in which to make the inference 'an doesn't fit with carrot' before the saccadic plans in place shift the eyes onwards? It is possible to cancel a saccadic plan but, at least in the EZ-Reader model, there

are time thresholds involved both in terms of the time taken to cancel a saccade, and the temporal closeness to the execution of the saccade after which it cannot be cancelled. Secondly, it may be that the availability of information provides a different behavioural pattern. In the ERP data, the participant is drip-fed words at a predefined speed, and has no access to information downstream until it is presented to them. In eye-tracking participants are aware of the noun, and able to fixate it whenever they choose, it may be that we see no main effect of predictability because participants do not feel the need to make a full inference of the article because the noun itself is available and provides not only information that their prediction has been violated, but also foveal information as to what the word actually is.

That said, in the probability of fixating the noun we do see a hint that some eye-movement control is being manipulated by prediction. This is not a main effect, but rather a marginal interaction between Prediction and Load such that there is a numerically higher proportion of fixations on the noun if the word is unpredicted compared to predicted (in the NoLoad condition). This increase in the likelihood of fixation when a prediction is violated is not significant however, and does not automatically indicate that the article has been skipped. Nevertheless, it indicates that fixation behaviour may be influenced by upcoming prediction violations, either because of para-foveal preview, or because the article indicated an incongruence. Perhaps a launch site analysis of where the saccades which contributed to the observed numerical increase were initiated from would reveal more about the source of the effect, but such an analysis is complex, post-hoc, and outwith the main focus of the thesis. As mentioned, the probability of fixation mitigation discussed here is the NoLoad condition of an interaction with Load. We will return to the implications of this interaction on the resource reduction aspect of the thesis shortly. First we will discuss the evidence observed pertaining to the Predictability by Frequency interactions which are much discussed in the literature.

We anticipated replicating the interactive effects found between Predictability and Frequency in the probability of fixating the critical noun observed by Rayner et al. (2004) and Hand et al. (2010). They found that only the easiest condition (Predictable and High Frequency) differed from the rest, and that it did so in a significantly reduced probability of fixating the noun. We did not replicate this effect. Conversely, Rayner et al. anticipated finding this interactive pattern in early reading times on the critical noun but did not, and

Hand et al. found mixed evidence for an interaction in early reading times only after adding, as an interacting factor, an index of the amount of para-foveal preview available to the reader. We found this interaction without conducting a launch site analysis, and interestingly we found it not only marginally in early reading measures (First Pass) but also significantly in several other reading and fixation measures. The most likely reason that we found this interaction in First Pass while others have not is that we found it in the combined region Article+Noun. This region provides a combination of foveal fixation durations on the a / an article (providing some confirmation that a prediction is correct) and the noun itself (providing full confirmation that the prediction is correct, and frequency information). Para-foveally, the a / an article also provides confirmation of predictions and frequency information of the noun. While Cloze tested, in the materials used by Rayner et al. and Hand et al. the word preceding the target noun was not an article, and did not agree / disagree with the noun in the way that our materials do. There are two ways in which this difference could cause this result. Firstly, the increase in total information provided by the combination of a). the validation of their prediction which readers gain on foveal fixations on the article, and b). the information available in the other studies (i.e. the para-foveal information gained on the word preceding the noun, and the foveal information gained on the noun), may be enough to allow the reader to process the easiest condition (predictable / high frequency) faster than the others. Secondly, the greater amount of information gained prior to the fixation of the noun, due to the validating / violating nature of the article, may allow the reader more time to conduct para-foveal processing on the noun, essentially leading to a greater level of pre-processing, and a reduction of time spent on the noun for the easiest condition. We would argue that the former explanation is more likely as should higher levels of pre-processing be taking place due to the influence of the article, we would expect to see skipping of the main noun in the easier conditions, as Rayner et al. and Hand et al. did.

Finally, we turn to the Prediction by Load interaction observed in the probability of fixating the noun. Note that this interaction does not involve frequency, and so is not pertinent to the discussion of the previous paragraph. This interaction is marginal and so should be considered with caution; however, assuming its validity for the time being, it appears that the reduction in resources we apply as our Load manipulation affects only one component of semantic prediction. It appears that the use of context to create predictions is unaffected, but when violated, those under Load are less able to notice, or less able to



adapt in a normal way. Normally we would anticipate an increased likelihood of fixating the noun if the article had violated a reader's expectation, and this is borne out by the NoLoad data (numerically, although not significantly). However, when under Load, we observe no statistical difference, and a (smaller) numerical difference in the opposite direction. Whether this is due to the loaded reader being unable to make use of the violating information provided by the article, or because in some way the adaptability of the saccade plan is slowed under Load so that, while the processing of violating information occurs, the reader's eye-movements remain unaffected, is unknown. Either way, the findings here appear to conflict with the predictions of the SR theory: The SR theory would predict mitigation in the creation of the initial prediction which is not observed here. Instead, the findings appear to support the SLIR theory as the initial contextual facilitation of predicting the noun is unaffected, and only when a violation occurs do we observe any difference predicated on Load. Although the SLIR theory does not specifically identify the violation of a prediction as being in either related to the interpretive or post-interpretive resource pools, it seems to correlate well with the processing of an ambiguity (in the sense of requiring more conscious attention) which is listed as under the purview of the post-interpretive resource pool.

Overall, it seems that the observation of prediction effects on the article that DeLong et al. (2005) report in their ERP experiment may be, as they suggest, the result of a neural mechanism which underlies language processing, but eye-tracking does not appear to be able to detect it very clearly: We see no main effect of skipping the noun predicated on the congruity of the article with the reader's prediction, but numerically the trends are in the anticipated direction. Whether we do not observe this effect significantly could be to do with several factors: A time scale issue (i.e. DeLong et al.'s participants had much longer to make use of the information provided by the article), a time constraint issue (i.e. saccade plans require time to make and cancel which may influence the opportunity for a reader to skip the noun even if they chose to), or a function of the availability of the upcoming noun (i.e. the additional information it provides above and beyond the indication of violation that the article delivers). A fourth option is that with more experimental power, or materials more tightly constructed to fit our testing population, the numerical trend indicating some predictive processing occurring on the article may become significant.

When considering the role of cognitive resources in this, it appears that the SLIR theory is more supported by the current data. If it were the case that with more power the numerical trends mentioned above became significant, the interpretation of how Load influences predictive processing would become less clear. On the one hand we would see the predicted pattern of skipping when not loaded and no difference in skipping when loaded, which would imply an overall degradation of the use of context in parsing congruent with the SR theory. However, the Predicted conditions are currently almost identical (with only the unpredicted conditions varying across Load) which, assuming that that remained the same, would indicate no difference in context use in parsing except when violations occur, which would still, regardless of the significance of the NoLoad skipping effect, fit better with the multiple resource theory of the SLIR. To conclude: The current data tentatively support the SLIR, and even if power issues were addressed and allowed the possibility of an SR interpretation, the SLIR interpretation would likely still be more plausible.

## Section 5: General Discussion

### Aim of Thesis

It was the intent of this thesis to examine the role of working memory resources in the processing of language. We chose three specific language challenges and investigated a). how contextual information was used to assist in their resolution, and b). how the use of such information was influenced by a reduction in available working memory resources. Specifically, we set out to test whether there is a single resource pool which feeds all language specific processes, or if those resources are fractionated into two (or more) pools and some language processes draw from different pools to others. We used an extrinsic memory load (which was titrated to each participant) to experimentally limit the amount of working memory resource available to that person. In this way, regardless of the normal capacity of the participant, all participants under Load experienced a reduction of resources to the same level. In the case of the single resource theory, this Loading reduced the overall capacity of the participant, while in the case of the multiple pools theory we suppressed only one of two pools of resources.

Critically, the two hypotheses associated with this debate (the Single Resource theory and the Separate Language Interpretation Resource theory) predict the same modularisation of processes in conditions of reduced resources (e.g. they would predict that the use of contextual information would be reduced or eliminated when resolving an ambiguity), but different patterns of results for when language proceeds normally (e.g. in the case of no ambiguity). Each of the language challenges we investigated (Syntactic ambiguity, Lexical ambiguity, and Semantic prediction) allowed for both normal and 'challenging' processing so we therefore had the opportunity to observe two interesting strains of information. Firstly, how is the processing of each challenge influenced by a reduction of working memory resources, and secondly, does the pattern observed fit with either the SR or SLIR theories.

## Overview and Implications of Empirical Studies

### Syntactic Ambiguity

In the syntactic ambiguity experiments we observed little evidence in support of a general degradation of contextual influence on ambiguity processing under conditions of resource reduction. Our principal observation in the NoLoad experiment (a Context by Ambiguity interaction such that fewer regressions out were made if the context supported the processing of the ambiguity, compared to when it didn't (and no difference for unambiguous words)) was not present in the Load experiment, but this was not statistically significant. There is a possibility that more experimental power might have brought this interaction into significance, but it is worth noting that in both experiments the pattern of results was the same, it was an overall suppression of the amount of regressions out across all conditions which led to the effect being non-significant in the Load experiment. It is therefore possible (although, in our opinion less likely) that an increase in power would highlight the similarities between the observations, rather than the differences.

In terms of normal (non-ambiguous processing) we see mixed evidence to impact the SR / SLIR question. On the one hand interactions between Context and Load (such that context is less influential under Load) support the general increase of modularity predicted by the SR. On the other hand, main effects of Context which do not interact with Load provide support for the SLIR.

This series of experiments proved inconclusive in terms of the SR / SLIR debate (although with a slight leaning towards the SR based on the presence of the Load by Context interactions mentioned above). It did, however, indicate that our Load manipulation was having a suppressive effect on the use of context in syntactic ambiguity resolution, although (likely due to a lack of experimental power) this turned out to be a non-significant difference. We reasoned that this lack of sufficient power might stem from a reduction in resources not being able to override the possibly privileged status of syntactic processing. One possibility is that resources can be flexibly applied as required; if the processing of a syntactic ambiguity is very high priority, then it is possible that resources (although already spread thinly across several processes) could be diverted from a less important process to assist with syntactic processing. This might leave a process without sufficient resources to

function, but failure of that process might not necessarily be observable in our experiments.

### Lexical Ambiguity

In the lexical ambiguity experiments we saw stronger evidence of contextual influence on ambiguity resolution being mitigated by Load. Our meta-analysis statistically confirmed the presence of a three-way interaction between Context, Ambiguity, and Load in the probability of fixating the critical region. The interaction patterned such that non-Loaded participants were less likely to fixate the critical region if the context supported one sense of the ambiguity compared to if it was neutral. No difference was observed if the critical word was not ambiguous. There was no evidence of this pattern in the fixation behaviour of Loaded participants. As it stands this effect indicates that an extrinsic memory load can directly impact the use of context in the resolution of lexical ambiguities. There is, however, an obvious concern raised by the nature of the effect. The interaction being observed in the probability of fixating the critical region necessarily implicates eye-movement control. The implications of this for eye-movement models (in terms of high-level factors like context impacting processing prior to fixation) have been discussed previously (sections 8.4, 10.4, 13.4); but here we highlight the concern that the interaction we observe being eliminated under Load may not be due to the impact of Load increasing modularity between processes (i.e. contextual influence on ambiguity resolution), but rather the inability of the reader to adapt their saccadic plans in time.

In either case, this interaction is not replicated in our within subjects design, although there are plausible task differences (to do with task switching and overall difficulty of the experiment) which might explain this.

In terms of the SR / SLIR debate, the interpretation of the three-way interaction discussed above is critical. If it is considered that it is either spurious (due to the failure to replicate it in the within subjects experiment) or due to Load influencing eye-movement control rather than Context by Ambiguity interactions, then it appears that our digit span manipulation had little or no effect. Note that there are other three way interactions which are discussed in the main text but not mentioned here that indicate some, if not consistent, influence of

Load on Context by Ambiguity interactions. If we consider the aforementioned interaction to be reliable, however, then it fits with the shared prediction of the SR and SLIR theories that modularity between processes will be increased as available resources are decreased.

In that case this series of experiments lends more credence to the SLIR theory as the effects of context in non-ambiguous processing remain under conditions of resource reduction, even as they are degraded in the presence of ambiguity.

### Semantic Predictability

In this experiment the main finding of note was a marginal interaction between Predictability and Load in the probability of fixating the head noun. The likelihood of making a fixation on the head noun was almost identical if the word was predicted by the context, regardless of Load. If it was not the predicted word, however, the likelihood of fixation was higher when participants were not under Load, compared to when they were Loaded. It appears that the creation of a prediction based on a preceding context is not influenced by a reduction in resources. Rather, it is the effect of having the prediction violated which is reduced when readers are under Load. This implies that at this level of parsing at least, contextual information is still applied, even when resources are limited. On the other hand, when resources are limited and a violation of the prediction occurs, the ability of a reader to react to the violation is reduced. Being observed in the probability of fixating the head noun, this interaction suffers the same uncertainty as to its origin as was noted in the discussion of the Lexical section above: it is possible that rather than being evidence of resource reduction influencing interactivity of language processes, it might simply reflect a reduction in the ability to quickly adapt saccadic plans.

Interestingly, though, if we do consider it to be the former option, then at the point the prediction is violated, the SLIR would predict a shift from normal processing drawing on the primary resource pool, to attentive processing drawing on the secondary pool. It is, in the view of the SLIR, this secondary pool of resources that our extrinsic memory load suppresses. Therefore, should we consider this to be an influence of Load on the use of Context in parsing, then this observation supports the SLIR theory.

## General Discussion

### Domain General or Domain Specific Resources

From this mixed data it is difficult to see a clear pattern of results. In its weakest form, the conclusion we would have to draw from the data presented here is that a Load of this kind does not significantly impact the use of context in language. This conclusion assumes that the interaction in the Syntactic section would not gain significance with the addition of more experimental power, the differences between the meta-analysis and the within subject experiment in the Lexical section is not down to task differences and is instead a spurious artefact of the between subjects design, and that the marginal interaction in the Prediction would not gain significance with more power, or is in fact an effect of Load on saccade planning rather than context use. We are concerned that to subscribe to this conclusion is to ignore a lot of evidence which, when combined, points to an interesting pattern of results.

Taking a more optimistic approach, a second interpretation would be that the results as a whole indicate a reduction in contextual use in processing under conditions of resources reduction, but that this reduction is variable in how it manifests across different processing challenges. Both syntactic and lexical ambiguity resolution appear to be affected in general, but in the syntax experiments we see mixed evidence for the SR and SLIR theories, while in the Lexical and Prediction sections we see evidence in support of the SLIR theory. The question is raised then, why do the experiments investigating syntactic ambiguity produce results more in keeping with the SR theory while the other two areas examined support the SLIR. There are three possible explanations to this. Firstly, we are seeing a distinction where none exists: the evidence in the syntactic experiments is mixed, and could be interpreted in favour of the SLIR if certain Context by Load interactions were explainable as related to some other phenomenon, or were spurious. The second is that only the effects of Load which are observed in the syntax section are truly effects of resource reduction on contextual processing, the ones observed in the Lexical and Prediction chapters are artefacts of eye-movement control programs being negatively affected by Load. This explanation is disputed by another interaction in the Lexical section (Context, Ambiguity, and Load interaction, First Pass, Spillover region – meta-analysis) but this interaction itself is

dubious as, while an interaction was found in the within subjects experiment which involved the same factors, in the same measure and region, the pattern was very different. Neither of these explanations is particularly satisfying. A third option is that there is something special about the processes involved in parsing relative clauses which necessitates a different approach to resource allocation. The SLIR theory hangs on the idea that attentive, more conscious processing is supplied by the post-interpretive resource pool, while more normal automatic processing is supplied by the interpretive pool. In the case of non-ambiguous processing, the SLIR argues that attentive processing should not occur and therefore only the interpretive resource should be tapped (so no effect of our load manipulation should be observed). Perhaps, relative clauses are infrequent, or unusual enough that some level of attention is paid to them, even when the relativizer is present to guide the parser. This would result in the tapping of the resource pool that our load manipulation is suppressing, and so implicates the SLIR theory rather than the SR theory. This would bring it in line with the other findings but is, admittedly, an untested and post-hoc explanation. Out of these three options, however, it does seem the most likely.

A fourth option, which is perhaps more appealing and less post-hoc, is that our memory load disproportionally influences certain processes. As noted in the literature review, digit span has been criticised as a measure of working memory as it only reflects storage capacity, and not the operational component of working memory. We reasoned that a). as we were trying to experimentally reduce working memory capacity rather than measure it, and b). given that the working memory system is considered to be a system in which resources are flexibly traded off between operations and storage; a task which occupied resources in storage would subsequently also limit operational functionality. It is possible, however, that our manipulation did not function in the way we anticipated, and as such might explain why our data is variable between experimental areas. If the digit span task is not tapping the resources fully (i.e. not particularly impacting the operational component) we might expect tasks which require more storage to suffer more than those that require less. This would fit with the observation that the data reported in the syntax section show some evidence of normal non-ambiguous processing being influenced by load, while the data from the Lexical and Prediction sections do not. The processing of relative clauses inherently has a higher memory cost than non-ambiguous lexical processing or the validation of non-violated predictions, and so might be more negatively influenced by a predominately storage based memory Load. If this explanation holds, then our discussion of



the SR / SLIR debate requires reframing. It would indicate that we have found fairly strong evidence to support the SR theory in the syntax section and that the memory requirements of the 'normal processing' in the other two experimental areas were not sufficient to observe the impact of our Load manipulation. We therefore gave credence to the SLIR theory when in fact it was a lack of sensitivity in our experimental manipulations which prohibited us from observing the influence of load on normal processing.

### Experimental reduction in resources vs. Low span readers

In the end of section discussion of the Syntax section (Chapter 6) we touched on the idea that being placed under conditions of resource reduction is not equivalent to having a naturally Low working memory span. This is a more complicated assertion than it might at first seem. Firstly, there are two schools of thought regarding the locus of the difference between High and Low span readers: the capacity constrained theory of Just & Carpenter (1992), and the subsequent Experiential account of MacDonald & Christiansen (2002). To address them in turn: A Low span reader in the capacity constrained theory simply has a lower overall working memory capacity than a High span reader. Working memory is defined as having both storage and operational components which are functionally divided by flexibly allocated resources. The theoretical grounding of our manipulation was that while digit span only taps storage resources, the lack of available storage resources would cause a trade-off such that operational resources would, when additional storage was required, be dynamically assigned as storage to make up the shortfall. Thus operational resources would be reduced as well. In terms of the capacity constrained model, this should cause our resource reduced participants to act as Low span readers. It is worth noting, however, that Low span readers as a group are only defined by the upper threshold of their working memory capacity. Individuals within that group may vary in how much lower than that threshold their maximum capacity actually is. By titrating the level of loading applied to each participant we aimed to reduce all participants to the same level which may have caused them to act differently to pre-defined Low Span groups.

It is possible, as we have noted that our experiments did not, in some circumstances, provide enough impetus for the operational / storage trade off mention above to occur. This represents a problem for comparison against Experiential accounts as they argue that

rather than a finite capacity limitation creating functional boundaries in working memory it is the efficiency of the operational processing which sets High and Low span readers apart. It is immediately apparent that a storage oriented resource limitation is not well suited to speaking to this account. However, in the case where our storage demands did exceed that of the remaining storage resource available, and operational processing was affected, we saw evidence that our resource reduced participants behaved differently to Low span readers in earlier studies (see Chapter 4, sections 0 & 4.4). This provides tentative support for at least some involvement of the efficiency of processes in functioning of working memory.

### Strategy and Task Demands

As we have noted throughout this thesis, there is a tendency for participants to skip more, read less thoroughly, and commit to re-reading less when they are under cognitive load. It has been noted that this might reflect task demands in which the desire to offload the digits (presumably before the memory traces decay) causes participants rush to the end of the sentence. While we do believe this possible if not controlled, we do not believe that our participants developed this as a conscious strategy. The reasons for this are two-fold. Firstly, we were careful to stress the importance of both tasks and measured performance on each with a view to discarding data which fell below predefined thresholds. Participants were made aware of this potential discarding, and reminded of it with feedback during the practice trials. Additionally, data from subjects who did fall below the predefined threshold on either task was discarded and another participant was run in their place. Secondly, while a little inconsistent and noisy, the data we report conforms in the main to predictions from the literature. It is possible, though, that the tasks demands and the desire to offload the digits did play a minor role in adding some noise to the data. For future work, we would stress the importance of limiting, as much as is possible, the scope for strategies such as this to influence results.

## Future Directions

One of the principal concerns about the methodology used in this thesis is the appropriateness of the digit recall task to properly reduce the availability of cognitive resources. While we maintain that the appropriate level and type of loading can be achieved using this methodology, it is clear that it produces noisy, inconsistent results unless controlled and applied very carefully. Other sources of loading exist (e.g. the interference effects shown by Gordon et al., 2002 and Fedorenko et al., 2006) , and while we are dubious about whether they truly tap cognitive resources, or if they instead impact processing on a different level (via interference of the retrieval of memory traces, for example) it might be interesting to examine the influence of memory load consisting of semantically related items, rather than unrelated digits. This would be applicable to our Lexical and Prediction materials, but less so to our syntactic materials.

Focusing more on the syntactic ambiguity experiments, it might be interesting to maintain the same loading system (now that we have a better understanding of the levels of difficulty required for effects to be observable) and vary the type of ambiguity or distance between the ambiguity onset and the disambiguating information. Some ambiguities require more storage resources to resolve than others, and the duration that a structural ambiguity is held in memory for would also increase the memory load. It would be interesting to demonstrate an increased modularity between processes (due to extrinsic memory loading) which varied in size depending on the type or duration of the ambiguity.

## Concluding Remarks

This thesis attempted to examine the role of working memory resources on influence of context on syntactic and lexical ambiguity resolution, and semantic prediction. The main question guiding the research was that of a Domain General or Domain Specific resources. Results were mixed and general conclusions hard to draw. We observed mixed evidence which, in sum, indicates a reduction of resources mitigating the use of context in the resolution of both syntactic and lexical ambiguities, and in adaptation when predictions were violated. Overall we conclude that our results are best understood with reference to

critical analysis of our working memory manipulation and how it relates to specific language processes. Syntactic ambiguity, with its greater reliance on storage resources, was influenced most by our mostly storage oriented load manipulation, and showed evidence to support Domain General resource theories. Lexical ambiguity and semantic prediction showed evidence to support Domain Specific resource theories, although it is possible, given that the storage requirements of these processes are low, that this reflects an insensitivity of these processes to the effects of load as applied by our manipulation.

In the process of this examination we have reported results of non-loaded processing for the three language processes we examined. We have replicated and extended findings in the literature, particularly in terms of Frequency and Predictability interactions, and have raised questions for models of eye-movement control regarding contextual processing in the para-fovea.

The research presented here contributes to the question of Domain General or Domain Specific resources. However, we feel that the main contribution of this thesis is in the examination and assessment of a method of cognitive loading across several different language processes. Overall we have found that the effects of a titrated extrinsic memory load based on the (storage oriented) digit span paradigm are hard to predict, noisy, and most importantly for further research: Variable across different language processes due to differing resource requirements.

## References

- Allport, D. (1979). Attention and Performance. In G. Claxton (Ed.), *New directions in cognitive psychology* (pp. 112–153). London: Routledge & Kegan Paul.
- Allport, D., Styles, E., & Hsieh, S. (1994). Shifting intentional set: Exploring the dynamic control of tasks. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance* (Vol. 15, pp. 421–452). Cambridge MA: The MIT Press.
- Altmann, G., Garnham, A., & Henstra, J. (1994). Effects of syntax in human sentence parsing: Evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 209–216.
- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, 30(3), 191–238.
- Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3), 247–264.
- Altmann, G., Garnham, A., & Dennis, Y. (1992). Avoiding the garden path: Eye movements in context. *Journal of Memory and Language*, 31(5), 685–712.
- Anderson, J. E., & Holcomb, P. J. (2005). An electrophysiological investigation of the effects of coreference on word repetition and synonymy. *Brain and Language*, 94(2), 200–216.
- Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In K. Spence (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (pp. 89–195). New York: Academic Press.
- Baayen, R. H. (2008). *Analysing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

- Baayen, R.H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Baddeley, A. (1998). The central executive: a concept and some misconceptions. *Journal of the International Neuropsychological Society : JINS*, 4(5), 523–526.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423.
- Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: evidence from task switching. *Journal of Experimental Psychology: General*, 130(4), 641–657.
- Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language*, 24(1), 119–131.
- Baddeley, A, & Warrington, E. K. (1970). Amnesia and the distinction between long-term and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 9(2), 176–189.
- Baddeley, A., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575–589.
- Baddeley, A. (2001). Is working memory still working. *American Psychologist*, 56(11), 851–864.
- Baddeley, A., Grant, S., Wight, E., & Thomson, N. (1975). Imagery and visual working memory. In P. Rabbitt & S. Dornic (Eds.), *Attention and Performance V* (pp. 205–217). New York: Academic Press.
- Baddeley, A., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed), *The Psychology of Learning and Motivation* 8, (pp.47–89). New York: Academic Press.
- Baddeley, A. (1992). Working Memory. *Science*, 255(5044), 556–559.

- Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology*, 49A(1), 5–28.
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews. Neuroscience*, 4(10), 829–839.
- Baddeley, A. (2010). Working memory. *Current Biology*, 20(4), 136–140.
- Balota, D., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17(3), 364–390.
- Barbizet, J. (1963). Defect of memorizing of hippocampal-mammillary origin: a review. *Journal of Neurology, Neurosurgery, and Psychiatry*, 26, 127–135.
- Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–73). Cambridge, UK: Cambridge University Press.
- Becker, J. T., & Morris, R. G. (1999). Working memory(s). *Brain and Cognition*, 41(1), 1–8.
- Bever, T. G. (1970). The Cognitive Basis for Linguistic Structures. In J. R. Hayes (Ed.), *Cognition and the Growth of Cognition* (pp. 279–362). New York, NY: Wiley.
- Binder, K., Duffy, S., & Rayner, K. (2001). The Effects of Thematic Fit and Discourse Context on Syntactic Ambiguity Resolution. *Journal of Memory and Language*, 44(2), 297–324.
- Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. (2011). Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, 26(3), 301–349.
- Bourke, P., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual task performance decrement. *Journal of Experimental Psychology: Human Experimental Psychology*, 49(3), 525–545.
- Britt, M. (1994). The interaction of referential ambiguity and argument structure in the parsing of prepositional phrases. *Journal of Memory and Language*, 33(2), 251–283.
- Broadbent, D. (1958). *Perception and Communication*. New York: Oxford University Press.

- Brooks, L. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 22(5), 349–368.
- Brooks, L. R. (1967). The suppression of visualization by reading. *The Quarterly Journal of Experimental Psychology*, 19(4), 289–299.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10(1), 12–21.
- Burgess, N., & Hitch, G. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language*, 31(4), 429–460.
- Caplan, D., & Waters, G. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 21(1), 77–126.
- Caplan, D., & Waters, G. (2002). Working memory and connectionist models of parsing: A reply to MacDonald and Christiansen (2002). *Psychological Review*, 109(1), 66–74.
- Chen, E., Gibson, E., & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, 52(1), 144–169.
- Cherry, E. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, 25(5), 975–979.
- Chomsky, N. (1957). *Syntactic Structures*. Den Haag: The Netherlands: Mouton.
- Clifton, C., & Ferreira, F. (1989). Ambiguity in Context. *Language and Cognitive Processes*, 4(3/4), 77–103.
- Clifton, Charles, & Staub, A. (2008). Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass*, 2(2), 234–250.
- Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, 4(3), 161–169.
- Conrad, R., & Hull, A. J. (1964). Information, Acoustic Confusion and Memory Span. *British Journal of Psychology*, 55(4), 429–432.



- Conway, A., Kane, M., Bunting, M., Hambrick, D., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5), 769–786.
- Copeland, D. E., & Radvansky, G. A. (2001). Phonological similarity in working memory. *Memory & Cognition*, 29(5), 774–776.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.
- Cowan, N., Nugent, L. D., Elliott, E. M., & Geer, T. (2000). Is there a temporal basis of the word length effect? A response to Service (1998). *The Quarterly Journal of Experimental Psychology: Section A*, 53(3), 647–660.
- Craik, F.I.M., Morris, R. G., & Gick, M. L. (1990). Adult age differences in working memory. In G. Vallar & T. Shallice (Eds.), *Neuropsychological impairments fo short-term memory* (pp. 247–267). Cambridge, UK: Cambridge University Press.
- Craik, F.I.M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684.
- Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural Language Parsing: Psychological, computational, and theoretical perspectives* (pp. 320–358). Cambridge, UK: Cambridge University Press.
- Crannell, C., & Parrish, J. (1957). A comparison of immediate memory span for digits, letters, and words. *Journal of Psychology*, 44(2), 319–328.
- Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.
- Daneman, M., & Merikle, P. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, 3(4), 422–433.

- Daneman, M. (1982). The measurement of reading comprehension: How not to trade construct validity for predictive power. *Intelligence*, 6(4), 331–345.
- Della Sala, S., Gray, C., Baddeley, A. Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwinding visuo-spatial memory. *Neuropsychologia*, 37(10), 1189–99.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121.
- Deutsch, J., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, 70(1), 80–90.
- Dubey, A., Sturt, P., Keller, F., & Stewart, O. W. (In Prep) High-cost referential support of relative clause ambiguities.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical Ambiguity and Fixation Times in Reading. *Journal of Memory and Language*, 27(4), 429-446.
- Ebbinghaus, H. (1913). Memory; a contribution to experimental psychology. In H. Ruger & C. Bussenius (Eds.), *Ebbinghaus, H.* New York: NY: Teachers College.
- Ehrlich, S. (1981). Children's word recognition in prose context. *Visible Language*, 15(3), 219–244.
- Ehrlich, S., & Rayner, K. (1981). Contextual Effects on Word Perception and Eye Movements during Reading. *Journal of Verbal Learning and Verbal Behavior*, 20(6), 641–655.
- Engbert, R., Longtin, A., & Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42(5), 621–636.
- Engbert, R., & Nuthmann, A. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review*, 112(4), 777–813.
- Ericsson, K., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102(2), 211–245.

- Farmer, T. A., Cargill, S. A., & Spivey, M. J. (2007). Gradiency and Visual Context in Syntactic Garden-Paths. *Journal of Memory and Language*, 57(4), 570–595.
- Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, 54(4), 541–553.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25(3), 348–368.
- Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: Bradford.
- Forster, K. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, 33(4), 465–495.
- Frazier, L. (1978). *On comprehending sentences: syntactic parsing strategies*. Doctoral Dissertation, Indiana University.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, 6(4), 291–325.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14(2), 178–210.
- Gardiner, J. M., Craik, F. I. M., & Birtwistle, J. (1972). Retrieval Cues and Release from Proactive Inhibition. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 778–783.
- Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-Load Interference in Syntactic Processing. *Psychological Science*, 13(5), 425–430.
- Grindrod, C. M., & Baum, S. R. (2003). Sensitivity to local sentence context information in lexical ambiguity resolution: Evidence from left- and right-hemisphere-damaged individuals. *Brain and Language*, 85(3), 503–523.

- Grindrod, C. M., & Baum, S. R. (2005). Hemispheric contributions to lexical ambiguity resolution in a discourse context: evidence from individuals with unilateral left and right hemisphere lesions. *Brain and Cognition*, 57(1), 70–83.
- Gunter, T., Jackson, J., & Mulder, G. (1992). An Electrophysiological Study of Semantic Processing in Young and Middle-Aged Academics. *Psychophysiology*, 29(1), 38–54.
- Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall: the roles of type of similarity, category cuing, and item recall. *Memory & Cognition*, 33(6), 1001–1016.
- Hale, J. (2001). A Probabilistic Earley Parser as a Psycholinguistic Model. *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*. (pp. 1-8). Association for Computational Linguistics.
- Hand, C. J., Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-predictability interaction in reading: it depends where you're coming from. *Journal of Experimental Psychology. Human Perception and Performance*, 36(5), 1294–1313.
- Hebb, D. (1949). *The organization of behavior*. New York: Wiley.
- Hegarty, M., Shah, P., & Miyake, A. (2000). Constraints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, 28(3), 376–385.
- Henson, R. (1998). Short-term memory for serial order: the Start-End Model. *Cognitive Psychology*, 36(2), 73–137.
- Henson, R., Norris, D. G., Page, M. P. A., & Baddeley, A. (1996). Unchained Memory: Error Patterns Rule out Chaining Models of Immediate Serial Recall. *The Quarterly Journal of Experimental Psychology A*, 49(1), 80–115.
- Honig, W. (1978). Studies of working memory in the pigeon. In S. Hulse, H. Fowler, & W. Honig (Eds.), *Cognitive Processes in Animal Behavior* (pp. 211–248). Hillsdale, NJ: Erlbaum.

- Ihara, A., Hayakawa, T., Wei, Q., Munetsuna, S., & Fujimaki, N. (2007). Lexical access and selection of contextually appropriate meaning for ambiguous words. *NeuroImage*, 38(3), 576–588.
- Inhoff, A., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: effects of word frequency. *Perception & Psychophysics*, 40(6), 431–439.
- Jackendoff, R. (2002). *Foundations of Language (Brain, Meaning, Grammar, Evolution)*. New York, NY: Oxford University Press.
- Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, 59(4), 434–446.
- Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? *Memory & Cognition*, 39(7), 1198–1210.
- James, W. (1890). *The Principles of Psychology, Vol. 1*. New York: Henry Holt and Co.
- Jersild, A. (1927). Mental set and shift. *Archives of Psychology*.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology. General*, 111(2), 228–238.
- Just, M. A., & Varma, S. (2002). A hybrid architecture for working memory: Reply to MacDonald and Christiansen (2002). *Psychological Review*, 109(1), 55–65.
- Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122–149.
- Just, M., Carpenter, P., & Keller, T. (1996). The capacity theory of comprehension: new frontiers of evidence and arguments. *Psychological Review*, 103(4), 773–780.
- Just, M., Carpenter, P., Keller, T., Eddy, W., & Thulborn, K. (1996). Brain activation modulated by sentence comprehension. *Science*, 274(5284), 114–116.
- Just, M., Carpenter, P., & Woolley, J. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology. General*, 111(2), 228–238.

- Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kamide, Y. (2008). Anticipatory Processes in Sentence Processing. *Language and Linguistics Compass*, 2(4), 647–670.
- Keller, F., Gunasekharan, S., Mayo, N., & Corley, M. (2009). Timing accuracy of Web experiments: a case study using the WebExp software package. *Behavior Research Methods*, 41(1), 1–12.
- King, J., & Just, M. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30(5), 580–602.
- Kroll, N. E., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, a L. (1970). Short-term memory while shadowing: recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, 85(2), 220–224.
- Kutas, M. (1987). Event-related brain potentials (ERPs) elicited during rapid serial visual presentation of congruous and incongruous sentences. *Current Trends in Event-Related Potential Research (EEG Suppl. 40)*, 406–411.
- Larsen, J. D., & Baddeley, A. (2003). Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: do they have a common source? *The Quarterly Journal of Experimental Psychology. Section A*, 56(8), 1249–1268.
- Lashley, K. (1951). The problem of serial order in behavior. In L. Jeffress (Ed.), *Cerebral Mechanisms in Behavior. The Hixon Symposium* (pp. 112–136). New York: John Wiley & Sons.
- Lee, C., & Estes, W. (1977). Order and Position in Primary Memory for Letter Strings. *Journal of Verbal Learning and Verbal Behavior*, 16(4), 395–418.
- Levy, B. A. (1971). Role of articulation in auditory and visual short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 10(2), 123–132.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1277.

- Lewandowsky, S., & Farrell, S. (2008a). Phonological similarity in serial recall: Constraints on theories of memory. *Journal of Memory and Language*, 58(2), 429–448.
- Lewandowsky, S., & Farrell, S. (2008b). Short-term memory: New data and a model. *Psychology of Learning and Motivation*, 49, 1–48.
- Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375–419.
- Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, 10(10), 447–454.
- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109(1), 35–54.
- MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language*, 32(5), 692–715.
- MacDonald, M. C., Just, M., & Carpenter, P. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24(1), 56–98.
- MacDonald, M. C., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703.
- Macnamara, B. N., Moore, A. B., & Conway, A. R. A. (2011). Phonological similarity effects in simple and complex span tasks. *Memory & Cognition*, 39(7), 1174–1186.
- Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research*, 1146, 115–127.
- McCann, R., & Johnston, J. (1992). Locus of the single-channel bottleneck in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, 18(2), 471–484.

- McClelland, J., & Rumelhart, D. (1986). Psychological and Biological Models. *Parallel distributed processing: explorations in the microstructure of cognition. Volume 2*. Cambridge, MA: MIT Press.
- McConkie, G., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception and Psychophysics*, 17(6), 578–586.
- McConkie, G., & Rayner, K. (1976). Asymmetry of the perceptual span in reading. *Bulletin of the Psychonomic Society*, 8, 365–368.
- McDonald, S., & Shillcock, R. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14(6), 648–652.
- McLaughlin, B. (1965). “Intentional” and “Incidental” Learning in Human Subjects: the Role of Instructions To Learn and Motivation. *Psychological Bulletin*, 63, 359–376.
- McLeod, P. (1977). A dual task response modality effect: Support for multiprocessor models of attention. *The Quarterly Journal of Experimental Psychology*, 29(4), 651–667.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the Influence of Thematic Fit (and Other Constraints) in On-line Sentence Comprehension. *Journal of Memory and Language*, 38(3), 283–312.
- Meiran, N. (1996). Reconfiguration of processing mode prior to task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1423–1442.
- Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 2(1), 1–21.
- Miller, G. (1956). The magical number seven, plus or minus two: Some limits on out capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Miller, G. A., & Isard, S. (1963). Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 2(3), 217–228.
- Miller, G., Galanter, E., & Pribram, K. (1960). *Plans and the structure of behavior*. New York: Holt.



- Milner, B. (1966). Amnesia following operation on the temporal lobes. In C. Whitty & O. Zangwill (Eds.), *Amnesia* (pp. 19–133). London: Butterworth.
- Mitchell, D., Corley, M., & Garnham, A. (1992). Effects of Context in Human Sentence Parsing: Evidence Against a Discourse-Based Proposal Mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(1), 69–88.
- Mitchell, D., & Green, D. (1978). The effects of context and content on immediate processing in reading. *The Quarterly Journal of Experimental Psychology*, 30(4), 609–636.
- Miyake, A., Just, M., & Carpenter, P. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neutral contexts. *Journal of Memory and Language*, 33(2), 175–202.
- Mohamed, M., & Clifton, C. (2011). Processing temporary syntactic ambiguity: The effect of contextual bias. *The Quarterly Journal of Experimental Psychology*, 64(9), 1797–1820.
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, 7(3), 134–140.
- Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, 11(1), 56–60.
- Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1353–1380.
- Nairne, J. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of psychology*, 53(1), 53–81.
- Navon, D. (1984). Resources - a theoretical soup stone? *Psychological Review*, 91(2), 216–234.
- Navon, D., & Gopher, D. (1979). On the Economy of the Human-Processing System, 86(3), 214–255.

- Nebel, K., Wiese, H., Stude, P., De Greiff, A., Diener, H., & Keidel, M. (2005). On the neural basis of focused and divided attention. *Cognitive Brain Research*, 25(3), 760–776.
- Neisser, U. (1967). *Cognitive Psychology* (p. 351). New York: Appleton-Century-Crofts.
- Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, 7(4), 480–494.
- Newell, A., & Simon, H. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- Norman, D. (1968). Toward a theory of memory and attention. *Psychological Review*, 75(6), 522–536.
- Norman, D. (1976). *Memory and Attention: An Introduction to Human Information Processing* (2nd ed., p. 254). New York: John Wiley & Sons.
- Norman, D., & Bobrow, D. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7(1), 44–64.
- Norman, D., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. Davidson, G. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (pp. 1–18). New York: Plenum.
- Nuthmann, A., & Engbert, R. (2009). Mindless reading revisited: an analysis based on the SWIFT model of eye-movement control. *Vision Research*, 49(3), 322–336.
- Olton, D., Becker, J., & Handelmann, G. (1980). Hippocampal function: Working memory or cognitive mapping? *Physiological Psychology*, 8(2), 239–246.
- Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, 9(3), 225–236.
- Page, M. P., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review*, 105(4), 761–781.

- Parkin, A. J. (1998). The central executive does not exist. *Journal of the International Neuropsychological Society*, 4(5), 518–522.
- Pashler, H. (1990). Do response modality effects support multiprocessor models of divided attention? *Journal of Experimental Psychology. Human Perception and Performance*, 16(4), 826–842.
- Pashler, H. (1994a). Graded capacity-sharing in dual-task interference? *Journal of Experimental Psychology. Human Perception and Performance*, 20(2), 330–342.
- Pashler, H. (1994b). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, 116(2), 220–244.
- Pearlmutter, N., & Macdonald, M. (1992). Plausability Effects and Syntactic Ambiguity Resolution. *Proceedings of the 14th annual conference of the Cognitive Science Society* (pp. 521–542).
- Pearlmutter, N., & MacDonald, M. (1995). Individual differences and probabilistic constraints in syntactic ambiguity resolution. *Journal of Memory and Language*, 34(4), 521–542.
- Perry, R., & Hodges, J. (1999). Attention and executive deficits in Alzheimer's disease. A critical review. *Brain*, 122(3), 383–404.
- Peterson, L., & Peterson, M. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58(3), 193–198.
- Posner, M. (1978). *Chronometric explorations of mind*. Hillside, NJ: Lawrence Erlbaum.
- Postman, L. (1964). Short-term memory and incidental learning. In Melton (Ed.), *Categories of Human Learning* (pp. 145–201).
- Princeton University "About WordNet." WordNet. Princeton University. 2010.  
<<http://wordnet.princeton.edu>>

- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 358–374.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14(3), 191–201.
- Rayner, K., & Frazier, L. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 15(5), 779–790.
- Rayner, K., Inhoff, A., Morrison, R., Slowiaczek, M., & Bertera, J. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology. Human Perception and Performance*, 7(1), 167–179.
- Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: implications for the E-Z Reader model. *Journal of Experimental Psychology. Human Perception and Performance*, 30(4), 720–732.
- Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading*. Englewood cliffs, NJ: Prentice Hall.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125–157.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (1999). Eye movement control in reading: accounting for initial fixation locations and refixations within the E-Z Reader model. *Vision Research*, 39(26), 4403–4411.

- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16(1), 1–21.
- Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience*, 139(1), 5–21.
- Richter, E. M., Engbert, R., & Kliegl, R. (2006). Current advances in SWIFT. *Cognitive Systems Research*, 7(1), 23–33.
- Rogers, R., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124(2), 207–231.
- Roland, D., Yun, H., Koenig, J.P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition*, 122(3), 267–79.
- Rönnerberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: a framework and a model. *International Journal of Audiology*, 42(S1), 68–76.
- Rudner, M., & Rönnerberg, J. (2008). The role of the episodic buffer in working memory for language processing. *Cognitive Processing*, 9(1), 19–28.
- Rumelhart, D., & McClelland, J. (1986). Foundations. *Parallel distributed processing: explorations in the microstructure of cognition. Volume 1*. Cambridge, MA: MIT Press.
- Ruthruff, E., Pashler, H. E., & Klaassen, A. (2001). Processing bottlenecks in dual-task performance: structural limitation or strategic postponement? *Psychonomic Bulletin & Review*, 8(1), 73–80.
- Saito, S. (1994). What effect can rhythmic finger tapping have on the phonological similarity effect? *Memory & Cognition*, 22(2), 181–187.
- Scarborough, D., Cortese, C., & Scarborough, H. (1977). Frequency and Repetition Effects in Lexical Memory. *Journal of Experimental Psychology. Human Perception and Performance*, 3(1), 1–17.

- Scarborough, D. L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory & Cognition*, 7(1), 3–12.
- Schwanenflugel, P., & Shoben, E. (1985). The Influence of Sentence Constraint on the Scope of Facilitation for Upcoming Words. *Journal of Memory and Language*, 24(2), 232–252.
- Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, 14(4), 489–537.
- Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *The Quarterly Journal of Experimental Psychology: Section A*, 51(2), 283–304.
- Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores : A neuropsychological study. *Quarterly Journal of Experimental Psychology*, 22(2), 261–273.
- Shiffrin, R., & Cook, J. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning and Verbal Behavior*, 17(2), 189–218.
- Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: a characterization of the mind's bottleneck. *PLoS Biology*, 3(2), e37.
- Slak, S. (1970). Phonemic recoding of digital information. *Journal of Experimental Psychology*, 86(3), 398–406.
- Slak, S. (1971). Long-term retention of random sequential digital information with the aid of phonemic recoding: A case report. *Perceptual and Motor Skills*, 33(2), 455–460.
- Spector, A., & Biederman, I. (1976). Mental set and mental shift revisited. *The American Journal of Psychology*, 89(4), 669–679.
- Spelke, E., Hirst, W., & Neisser, U. (1976). Skills of divided attention. *Cognition*, 4(3), 215–230.

- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 24(6), 1521–1543.
- Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science*, 253(5026), 1380–1386.
- Sternberg, R. (2003). *Cognitive Psychology* (3rd ed., p. 597). Thomson Wadsworth.
- Sturt, P., & Crocker, M. (1996). Monotonic syntactic processing: A cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, 11(5), 449–494.
- Sudevan, P., & Taylor, D. A. (1987). The cueing and priming of cognitive operations. *Journal of Experimental Psychology. Human Perception and Performance*, 13(1), 89–103.
- Swaab, T., Brown, C., & Hagoort, P. (2003). Understanding words in sentence contexts: The time course of ambiguity resolution. *Brain and Language*, 86(2), 326–343.
- Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18(6), 645–659.
- The British National Corpus*, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>
- Tabossi, P., Spivey-Knowlton, M. J., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 589–615). Cambridge, MA: MIT Press.
- Tabossi, Patrizia, Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, 49(2-3), 161–167.
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634.

- Tehan, G., Hendry, L., & Kocinski, D. (2001). Word length and phonological similarity effects in simple, complex, and delayed serial recall tasks: Implications for working memory. *Memory, 9*(4-6), 333–348.
- Telford, C. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14*(1), 1–36.
- Tombu, M., & Jolicoeur, P. (2003). A central capacity sharing model of dual-task performance. *Journal of Experimental Psychology: Human Perception and Performance, 29*(1), 3–18.
- Traxler, M., & Foss, D. (2000). Effects of sentence constraint on priming in natural language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1266–1282.
- Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct Attachment Is Not a Form of Lexical Ambiguity Resolution. *Journal of Memory and Language, 39*(4), 558–592.
- Traxler, M. J., & Tooley, K. M. (2007). Lexical mediation and context effects in sentence processing. *Brain research, 1146*, 59–74.
- Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*(4), 242–248.
- Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior, 3*(6), 449–459.
- Trueswell, J., Tanenhaus, M., & Garnsey, S. (1994). Semantic Influence on Syntactic Processing: Use of Thematic Information in Syntactic Disambiguation. *Journal of Memory and Language, 33*(3), 285–318.
- Tulving, E. (1989). Memory: Performance, knowledge, and experience. *European Journal of Cognitive Psychology, 1*(1), 3–26.



- Tulving, E., & Gold, C. (1963). Stimulus information and contextual information as determinants of tachistoscopic recognition of words. *Journal of Experimental Psychology*, 66(4), 319–327.
- Vallar, G., & Baddeley, A. (1982). Short-term forgetting and the articulatory loop. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 34(1), 53–60.
- Van Berkum, J. J. A, Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: evidence from ERPs and reading times. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 31(3), 443–467.
- Van Dyke, J. A, & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language*, 55(2), 157–166.
- Van Gompel, R. P. G., Pickering, M. J., Pearson, J., & Liversedge, S. P. (2005). Evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, 52(2), 284–307.
- Van Gompel, R. P. G., Pickering, M. J., & Traxler, M. J. (2001). Reanalysis in Sentence Processing: Evidence against Current Constraint-Based and Two-Stage Models. *Journal of Memory and Language*, 45(2), 225–258.
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176–190.
- Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology*, 95(1), 57–79.
- Vandierendonck, A., Vooght, G. De, & Van der Goten, K. (1998). Interfering with the Central Executive by Means of a Random Interval Repetition Task. *The Quarterly Journal of Experimental Psychology: Section A*, 51(1), 197–218.

- Vos, S., & Friederici, A. (2003). Intersentential syntactic context effects on comprehension: The role of working memory. *Cognitive Brain Research*, 16(1), 111–122.
- Wanner, E., & Maratsos, M. (1978). An ATN approach to comprehension. In M. M. Halle, G. Miller, & J. Bresnan (Eds.), *Linguistic Theory and Psychological Reality*. MIT Press.
- Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain: a Journal of Neurology*, 92(4), 885–896.
- Waters, G., & Caplan, D. (1996). The capacity theory of sentence comprehension: critique of Just and Carpenter (1992). *Psychological Review*, 103(4), 761–772.
- Welford, A. (1952). The “psychological refractory period” and the timing of high-speed performance—a review and a theory. *British Journal of Psychology. General Section*, 43(2), 2–19.
- West, R., & Stanovich, K. (1982). Source of Inhibition in Experiments on the Effect of Sentence Context on Word Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(5), 385–399.
- Wicha, N., Bates, E., Moreno, E., & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346(3), 165–168.
- Wicha, N., Moreno, E., & Kutas, M. (2004). Anticipating words and their gender: an event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience*, 16(7), 1272–1288.
- Wickelgren, W. (1965). Short-term memory for phonemically similar lists. *The American Journal of Psychology*, 78(4), 567–574.
- Wickelgren, W. A. (1968). Sparing of short-term memory in an amnesic patient: Implications for strength theory of memory. *Neuropsychologia*, 6(3), 235–244.

## Appendix 1: Materials for Syntactic Ambiguity Experiments

Below are the materials for the Syntactic Ambiguity Experiments. We present an example of each context condition, followed by one example of the target and lead-out sentences. The underlined sections highlight the differences between the context conditions (with the non-italic ones being a non-supportive context (presented first), and those in italic being a supportive context (presented second)) and the parentheses indicate the conditional presence / absence of the relativizer in the target sentence.

1.

Yesterday at the nursery, two newborns suffered asthma attacks. The paramedics arrived and carried one newborn to the intensive care section, but the other had recovered by the time they arrived.

Yesterday at the nursery, *a newborn and a carer* suffered asthma attacks. The paramedics arrived and carried one newborn to the intensive care section, but the *carer* had recovered by the time they arrived.

The newborn (who was) carried by the paramedics soon recovered, though. A few breaths from the respiration machine helped him a lot.

2.

Two customers ordered drinks at the bar. The bartender was preoccupied and only served drinks for one of the customers.

*A customer and an off-duty employee* ordered drinks at the bar. The bartender was preoccupied and only served drinks for *the customer*.

The customer (who was) served by the bartender was unhappy with her cocktail, though. She thought it was too sweet and didn't have enough alcohol.

3.

The newspaper was considering two candidates to fill an empty position. The editor only interviewed one of the candidates because he thought the other one didn't have a strong CV.

The editor was considering if he should hire his nephew or an external candidate who applied for the same job. He only needed to interview the candidate as he knew his nephew quite well.

The candidate (who was) interviewed by the editor was unfit for the job. Although his resume looked promising, the editor found him to be immature and egotistical, and decided to hire his nephew.

4.

A detective recently investigated two well-known crooks who were suspected of robbing a bank. The detective located and arrested one crook, but the other crook is still at large.

A detective recently investigated a crook and a bank employee who were suspected of robbing a bank. The detective arrested the crook, but didn't have enough evidence to arrest the employee.

The crook (who was) arrested by the detective was identified by a bank customer. Although the detective knew someone else was involved, there was only enough evidence for one arrest.

5.

The Smith twins, still only babies, are no strangers to excitement. Only last week, a freak storm caused a fire which frightened one of the babies, but not the other.

The Smith family's baby and 7-year-old are no strangers to excitement. Only last week, a freak storm caused a fire which frightened the baby, but not the 7-year-old.

The baby (who was) frightened by the flames was crawling away as quickly as she could. Her mother picked her up and tried to calm her down.

6.

Fred and his teenage son spotted two deer while they were hunting. His son only shot down one of the deer even though he fired at both of them.

Fred and his teenage son spotted a deer and a rabbit while they were hunting. His son only shot down the deer even though he fired at both of them.

The deer (who was) shot down by the teenager was only used as a trophy. He was the biggest stag killed that year.

7.

Two suspects are believed to be the masterminds behind the drug trade in the small town of Arcata, California. An anonymous tip-off resulted in the FBI interrogating one suspect, but they are still gathering evidence against the other suspect.

In their latest drug case, the FBI had a well-placed informant who provided evidence against a suspect. The FBI only needed to interrogate the suspect as the informant had been wearing a listening device.

The suspect (who was) interrogated by the FBI was later found guilty of drug trafficking. He had been obtaining cocaine from a narcotics officer.

8.

After several days in the deep jungle, a group of mercenaries encountered two pigs in a small settlement. They stole and slaughtered one of the pigs and left the other in the settlement.

After several days in the deep jungle, a group of mercenaries encountered a pig and a sheep in a small settlement. They stole and slaughtered the pig and left the sheep in the settlement.

The pig (who was) slaughtered by the mercenaries was completely eaten in one meal. It was the first time they had eaten since being in the jungle.

9.

The soldiers stormed the bunker, which was occupied by two hostages. The soldiers rescued one of the hostages, but enemy agents fled with the other hostage.

The soldiers stormed the bunker, which was occupied by an enemy agent and his hostage. The soldiers rescued the hostage, and killed the enemy agent.

The hostage (who was) rescued by the soldiers was happy to be a free man. He had been held for almost a year.

10.

The bus company's trustees hosted a lunch where they heard some insightful comments from two guests. They had specifically invited one guest, whereas the other guest, a German transportation expert, had arrived unannounced.

The bus company's trustees hosted a lunch where they heard some insightful comments from a quest and from one of their senior bus drivers. The quest they had specifically invited, whereas the driver had arrived unannounced.

The guest (who was) invited by the trustees gave a speech at the lunch. She presented her firm's proposal concerning new methods of public transport.

11.

A therapist was having problems with two patients who were having difficulty opening up about their pasts. She tried some unconventional treatments: she hypnotised one patient and used art therapy on the other.

A therapist was having problems with her receptionist, who wanted a large payrise, and a troubled patient, who couldn't talk about his childhood. Finally, she ended up firing the receptionist and treating the patient with hypnosis.

The patient (who was) hypnotised by the therapist cried like a baby. Clearly, the abuse he had suffered as a child was still with him.

12.

Two terrorists were involved in a recent airline bomb threat. They were captured, and the CIA tortured one of the terrorists and executed the other one.

A terrorist and a respected Saudi prince were involved in a recent bomb threat. Both were extradited to Yemen, where the CIA tortured the terrorist, and executed the prince.

The terrorist (who was) tortured by the CIA had secret information about his superiors. That was the only reason they kept him alive.

13.

Two witnesses came forward with information about a recent shooting in Glasgow. The police only questioned one witness as they knew that the other was unreliable.

A witness and a friend of the suspect came to the police with information. The police only questioned the witness as they knew the friend would just cover for his mate.

The witness (who was) questioned by the police was being overly defensive, though. The sergeant's doubts about his honesty increased throughout the interview.

14.

Two patients were given a treatment for chemical inhalation after they inhaled toxins in a chemical fire. The treatment cured one of the patients, but the other, unfortunately, did not make it.

A patient and a nurse were given a treatment for chemical inhalation after a canister broke in the hospital. The treatment cured the patient, despite his other health problems, but the nurse did not make it.

The patient (who was) cured by the treatment had been diagnosed as terminally ill. Needless to say, he was quite ecstatic about the outcome.

15.

The insurgents claimed that two martyrs provided inspiration to their members. The government had executed one martyr and the other martyr died in a suicide bombing.

The insurgents claimed that a martyr and their charismatic leader provided inspiration to their members. The government had executed the martyr and the leader's speech was all over the news.

The martyr (who was) executed by the government was caught while watching a movie. His hate for the government could not keep him away from kung-fu action movies.

16.

Two suspects were being tried for the gruesome murder of a bus driver. When one of the suspects was testifying, the barrister accused him of lying under oath as the other suspect watched in fear.

That day, a witness and a suspect had testified about the gruesome murder of a bus driver. When the suspect was testifying, the barrister accused him of lying under oath as the witness watched with interest.

The suspect (who was) accused by the lawyer was trying to cover for his wife. Although he knew she had committed the murder, he felt he couldn't live without her.

17.

A comedian was visiting town, and played at two shows, one on a Thursday and one on a Friday. The audience on Friday found him hilarious, while the audience on Thursday heckled more than they laughed.

A comedian was visiting town, and the club's staff had a different reaction from the audience. The audience found him hilarious, while the staff were extremely annoyed with his off-stage behaviour.

The audience (who was) entertained by the comedian had laughed heartily at his every word. He was good at finding humour in the most mundane activities.



18.

When two celebrities entered the posh club, the paparazzi raised their cameras.

Fortunately for the other celebrity, the photographers only recognised one of the two.

When a celebrity and a businessman entered the posh club, the paparazzi raised their cameras. Fortunately for the businessman, the photographers only recognised the celebrity.

The celebrity (who was) recognised by the photographers tried her best to get away before they took any pictures. She had been hounded by the press for weeks.

19.

The tribesmen had roasted two rabbits that afternoon. They devoured one of the rabbits, and offered the other as a sacrifice.

The tribesmen roasted a rabbit that afternoon, and placed a ceremonial wreath near the fire. They devoured the rabbit after bowing toward the wreath.

The rabbit (who was) devoured by the tribesmen had been roasting over the coals all afternoon. The meal was the centrepiece of their ritual to celebrate the new moon.

20.

Jimmy, the class bully, liked to pick on two wimps from his school. Just yesterday, Jimmy kicked one of the wimps and yelled at the other.

Jimmy, the class bully, liked to pick on a wimp and a new teacher at his school. Just yesterday, Jimmy kicked the wimp and yelled at the teacher.

The wimp (who was) kicked by the bully had tried to stand up to his tormentor. His determination to get back at the bully grew stronger each day.

21.

A young mother became exasperated with her two children yesterday. She punished one of her children and ignored the other child.

A young mother became exasperated with her child and her husband yesterday. She punished her child and ignored her husband.

The child (who was) punished by her mother had been trying to steal sweets. Sometimes, family life can be quite annoying.

22.

The peasants believed that two goddesses controlled their destiny. They worshipped one of the goddesses and feared the other.

The peasants believed that a goddess and a demon controlled their destiny. They worshipped the goddess and feared the demon.

The goddess (who was) worshipped by her followers was ignorant of their strife. She often wished they would leave her alone and just go away.

23.

Two freshers earned themselves a notorious reputation on campus. The student affairs committee lectured one of the freshers, while the other fresher's escapades became legendary among his peers.

A fresher and a new lecturer earned themselves a notorious reputation on campus. The student affairs committee lectured the fresher, while the lecturer's escapades became legendary among his peers.

The fresher (who was) lectured by the committee was told that he would be suspended. He had been caught vomiting in the residence halls on three separate occasions.

24.

The worried parents told the firemen that they had two infants. One infant was already safely outside, so the fireman rushed into the flames, where he saw the trapped infant, and lifted her up.

The worried parents told the firemen that they had a dog and an infant. The dog was already safely outside, so the fireman rushed into the flames, where he saw the trapped infant, and lifted her up.

The infant (who was) lifted by the fireman was unable to breathe. Once outside, though, they saw that she was a bit shaken, but otherwise fine.

25.

Two rare Arabian mice had escaped from their cage at the zoo. The zookeeper found one mouse sitting just outside the cage, but had to chase the other rodent around the compound.

A rare Arabian mouse and a colourful bird had escaped from their cages at the zoo. The zookeeper found the bird perched just outside its cage, but had to chase the mouse around the compound.

The mouse (who was) chased by the zookeeper had escaped once before. Its teeth were apparently stronger than the wire in its cage.

26.

The schools inspector was determined to punish two troublesome students. He dismissed one of the students, and harshly reprimanded the other student.

The schools inspector was determined to punish a troublesome student and a lazy teacher. He dismissed the student, and harshly reprimanded the teacher in front of his class.

The student (who was) dismissed by the schools inspector was happy to be leaving the room. He was tired of being accused of being the school's main drug source, even though he was.

27.

Two gangsters dominated the local crime scene. The government investigated one gangster because the other gangster was still unknown to public officials.

A gangster and a respected business man dominated the local crime scene. The government investigated the gangster because businessman's criminal connections were unknown to public officials.

The gangster (who was) investigated by the government was accused of assassinating his father. Assassination is a common end for the head of a crime family.

28.

After a fight at a rowdy pub, the police decided to charge two people. One defendant was examined by a police doctor while the other defendant gave an oral statement.

After a fight at a rowdy pub, the police arrested a defendant based on evidence from an eye-witness. The defendant was examined by a police doctor while the eye-witness gave an oral statement.

The defendant (who was) examined by the police medical examiner had been badly injured. Bar fights were common and the medical examiner did this often.

## Appendix 2: Materials for Lexical Ambiguity Experiments

Below are the materials used in the Lexical Ambiguity experiments. The ambiguous word is presented above the set of materials with the frequency and length matched control in parentheses. The two Neutral context conditions are presented first, followed by the two Supportive contexts. Two versions of the Neutral / Supportive contexts were created for each item (for counterbalancing reasons) and are presented here as two separate sets. For the within subjects experiment, only items containing an ambiguous word were presented.

Set 1:

### Appendix (fragment)

The woman was unhappy with the way things were going.  
It was no good, the fragment would have to be removed.

The woman was unhappy with the way things were going.  
It was no good, the appendix would have to be removed.

The woman was complaining of abdominal pain.  
It was no good, the fragment would have to be removed.

The woman was complaining of abdominal pain.  
It was no good, the appendix would have to be removed.

### Bark (pest)

The boys were playing in the park.  
The boys agreed that the pest was very unpleasant.

The boys were playing in the park.  
The boys agreed that the bark was very unpleasant.

The boys were climbing trees in the woods.  
The boys agreed that the pest was very unpleasant.

The boys were climbing trees in the woods.  
The boys agreed that the bark was very unpleasant.

### Cabinet (content)

Unlike the previous one, this decision was proving very tricky.  
He was aware that choosing the content was very important.

Unlike the previous one, this decision was proving very tricky.  
He was aware that choosing the cabinet was very important.

The Prime Minister had a created a shortlist.  
He was aware that choosing the content was very important.

The Prime Minister had a created a shortlist.  
He was aware that choosing the cabinet was very important.

### Calves (joints)

John had been very busy but could not relax.  
There was a problem with his joints that needed addressing.

John had been very busy but could not relax.  
There was a problem with his calves that needed addressing.

The man had been running for nearly an hour but had to stop.  
There was a problem with his joints that needed addressing.

The man had been running for nearly an hour but had to stop.  
There was a problem with his calves that needed addressing.

### Change (policy)

Jane was furious at the situation.  
It soon became clear that the policy was a problem.

Jane was furious at the situation.  
It soon became clear that the change was a problem.

Jane was upset that they weren't allowed to fix the broken vending machine.  
It soon became clear that the policy was a problem.

Jane was upset that they weren't allowed to fix the broken vending machine.  
It soon became clear that the change was a problem.

#### Chest (clock)

It had not been a good day for the man in the green hat.  
The man was upset that his clock had been badly burned.

It had not been a good day for the man in the green hat.  
The man was upset that his chest had been badly burned.

When he arrived he could see that the furniture had been damaged by the fire.  
The man was upset that his clock had been badly burned.

When he arrived he could see that the furniture had been damaged by the fire.  
The man was upset that his chest had been badly burned.

#### Class (event)

They were still recovering from the shock.  
Moving so quickly between each event had been an enlightening experience.

They were still recovering from the shock.  
Moving so quickly between each class had been an enlightening experience.

The night began in a sleazy pub and ended in a ballroom at the palace.  
Moving so quickly between each event had been an enlightening experience.

The night began in a sleazy pub and ended in a ballroom at the palace.  
Moving so quickly between each class had been an enlightening experience.

#### Coach (actor)

They found themselves with a little money left over.  
Unsurprisingly, the new actor was much better than the previous one.

They found themselves with a little money left over.  
Unsurprisingly, the new coach was much better than the previous one.

They had hired some fresh talent in order to succeed.  
Unsurprisingly, the new actor was much better than the previous one.

They had hired some fresh talent in order to succeed.  
Unsurprisingly, the new coach was much better than the previous one.

#### Fan (cat)

The people nearby were beginning to show signs of irritation.  
Everyone was sick of the cat making so much noise.

The people nearby were beginning to show signs of irritation.  
Everyone was sick of the fan making so much noise.

In the end they had to kick him out of the house.  
Everyone was sick of the cat making so much noise.

In the end they had to kick him out of the house.  
Everyone was sick of the fan making so much noise.

#### Match (radio)

Silence descended on the group as the realisation set in.  
Everyone was surprised that the radio had lasted as long as it did.

Silence descended on the group as the realisation set in.  
Everyone was surprised that the match had lasted as long as it did.

Lost in the woods with darkness closing they tried to start a fire.  
Everyone was surprised that the radio had lasted as long as it did.

Lost in the woods with darkness closing they tried to start a fire.  
Everyone was surprised that the match had lasted as long as it did.

#### Matter (report)

It took nearly four hours of discussion and a whole pack of coffee.  
In the end, they had all agreed that the report was very important.

It took nearly four hours of discussion and a whole pack of coffee.  
In the end, they had all agreed that the matter was very important.

The health care policy regarding depression had been in place for a week.  
In the end, they had all agreed that the report was very important.

The health care policy regarding depression had been in place for a week.  
In the end, they had all agreed that the matter was very important.



#### Mold (rack)

The rules in the building were not enforced very well.  
Despite regulations, it was not unusual for the rack to be left out overnight.

The rules in the building were not enforced very well.  
Despite regulations, it was not unusual for the mold to be left out overnight.

It was odd that the jelly set equally well at room temperature as in the fridge.  
Despite regulations, it was not unusual for the rack to be left out overnight.

It was odd that the jelly set equally well at room temperature as in the fridge.  
Despite regulations, it was not unusual for the mold to be left out overnight.

#### Nails (bolts)

The sun was setting when he finally arrived at the house.  
It became apparent that he had lost a few bolts on the way.

The sun was setting when he finally arrived at the house.  
It became apparent that he had lost a few nails on the way.

It was a big job and the builder had gone to get some special equipment.  
It became apparent that he had lost a few bolts on the way.

It was a big job and the builder had gone to get some special equipment.  
It became apparent that he had lost a few nails on the way.

#### Panel (crowd)

In situations like these it often paid to be observant.  
The young man with glasses stood as close to the crowd as he could.

In situations like these it often paid to be observant.  
The young man with glasses stood as close to the panel as he could.

The contestants were nervously awaiting comments on their performances.  
The young man with glasses stood as close to the crowd as he could.

The contestants were nervously awaiting comments on their performances.  
The young man with glasses stood as close to the panel as he could.

#### Pupil (mouth)

Once again, he was having difficulty.  
There was something wrong with his mouth that needed checking out.

Once again, he was having difficulty.  
There was something wrong with his pupil that needed checking out.

The accountant was on his way to the hospital to see a specialist.  
There was something wrong with his mouth that needed checking out.

The accountant was on his way to the hospital to see a specialist.  
There was something wrong with his pupil that needed checking out.

#### Ruler (graph)

There were subtle discrepancies, observable only if you were paying attention.  
Upon inspection it turned out that the graph wasn't entirely accurate.

There were subtle discrepancies, observable only if you were paying attention.  
Upon inspection it turned out that the ruler wasn't entirely accurate.

Even the military advisors could tell when they were being lied to.  
Upon inspection it turned out that the graph wasn't entirely accurate.

Even the military advisors could tell when they were being lied to.  
Upon inspection it turned out that the ruler wasn't entirely accurate.

#### Sense (light)

John had been struggling to make progress but eventually made a breakthrough.  
He realised he'd been employing the wrong kind of light for the task at hand.

John had been struggling to make progress but eventually made a breakthrough.  
He realised he'd been employing the wrong kind of sense for the task at hand.

John had tried to think about it logically but in the end just went with his intuition.  
He realised he'd been employing the wrong kind of light for the task at hand.

John had tried to think about it logically but in the end just went with his intuition.  
He realised he'd been employing the wrong kind of sense for the task at hand.

Yard (gate)

Apparently, some small towns have interesting bylaws.  
For example, there has to be a gate between two houses.

Apparently, some small towns have interesting bylaws.  
For example, there has to be a yard between two houses.

The planning regulations are in imperial measurements and are oddly specific.  
For example, there has to be a gate between two houses.

The planning regulations were in imperial measurements and oddly specific.  
For example, there has to be a yard between two houses.

Set 2:

Appendix (fragment)

They were all a bit concerned about the situation.  
Eventually they decided that the fragment wasn't a problem.

They were all a bit concerned about the situation.  
Eventually they decided that the appendix wasn't a problem.

It was the third time she had been in hospital that month.  
Eventually they decided that the fragment wasn't a problem.

It was the third time she had been in hospital that month.  
Eventually they decided that the appendix wasn't a problem.

Bark (pest)

It was quiet and peaceful as they walked through the countryside.  
Mary was the only one who didn't find the pest extremely annoying.

It was quiet and peaceful as they walked through the countryside.  
Mary was the only one who didn't find the bark extremely annoying.

Mary was making a swing in the tree at the bottom of the garden.  
She was the only one who didn't find the pest extremely annoying.

Mary was making a swing in the tree at the bottom of the garden.  
She was the only one who didn't find the bark extremely annoying.

### Cabinet (content)

It was the final thing they had to deal with and they were all tired.  
Even though the content was complete, they still couldn't make a decision.

It was the final thing they had to deal with and they were all tired.  
Even though the cabinet was complete, they still couldn't make a decision.

The politicians were pleased with their selections but the meeting wasn't over.  
Even though the content was complete, they still couldn't make a decision.

The politicians were pleased with their selections but the meeting wasn't over.  
Even though the cabinet was complete, they still couldn't make a decision.

### Calves (joints)

It was a sad day for all but Fred was particularly distraught.  
He had been informed that the joints were the often the first to go.

It was a sad day for all but Fred was particularly distraught.  
He had been informed that the calves were the often the first to go.

The runner was not surprised when he injured himself.  
He had been informed that the joints were the often the first to go.

The runner was not surprised when he injured himself.  
He had been informed that the calves were the often the first to go.

### Change (policy)

It had been another good day at work.  
It was clear that the policy made the barmaid happy.

It had been another good day at work.  
It was clear that the change made the barmaid happy.

The pub owner had always made sure that all the tips went straight to the staff.  
It was clear that the policy made the barmaid happy.

The pub owner had always made sure that all the tips went straight to the staff.  
It was clear that the change made the barmaid happy.

#### Chest (clock)

The man was feeling a bit depressed.  
He was very happy with his clock but nothing else.

The man was feeling a bit depressed.  
He was very happy with his chest but nothing else.

The King had commissioned new furniture for the palace.  
He was very happy with his clock but nothing else.

The King had commissioned new furniture for the palace.  
He was very happy with his chest but nothing else.

#### Class (event)

Trouble always seemed to follow him around.  
As usual the reputation of his event was the major problem.

Trouble always seemed to follow him around.  
As usual the reputation of his class was the major problem.

He had always felt that society had looked down on him.  
As usual the reputation of his event was the major problem.

He had always felt that society had looked down on him.  
As usual the reputation of his class was the major problem.

#### Coach (actor)

Lately things had not been particularly restful.  
It became apparent that the new actor was much louder than the old one.

Lately things had not been particularly restful.  
It became apparent that the new coach was much louder than the old one

They all got extremely drunk in a karaoke bar to celebrate their success.  
It became apparent that the new actor was much louder than the old one.

They all got extremely drunk in a karaoke bar to celebrate their success.  
It became apparent that the new coach was much louder than the old one.

#### Fan (cat)

This was a blemish on an otherwise good day.  
The man standing next to the cat looked very uncomfortable.

This was a blemish on an otherwise good day.  
The man standing next to the fan looked very uncomfortable.

Their team was losing and the supporters in the pub were taking it badly.  
The man standing next to the cat looked very uncomfortable.

Their team was losing and the supporters in the pub were taking it badly.  
The man standing next to the fan looked very uncomfortable.

#### Match (radio)

They had heard many good reports so decided to give it a try.  
After giving it a chance, the boys found the radio to be disappointing.

They had heard many good reports so decided to give it a try.  
After giving it a chance, the boys found the match to be disappointing.

The shopkeeper told them that it was highly water resistant.  
After giving it a chance, the boys found the radio to be disappointing.

The shopkeeper told them that it was highly water resistant.  
After giving it a chance, the boys found the match to be disappointing.

#### Matter (report)

As he walked into work he knew something big had happened.  
Sadly, it turned out that the report was being discussed a lot.

As he walked into work he knew something big had happened.  
Sadly, it turned out that the matter was being discussed a lot.

They had tried to keep the policy decision out of the press.  
Sadly, it turned out that the report was being discussed a lot.

They had tried to keep the policy decision out of the press.  
Sadly, it turned out that the matter was being discussed a lot.

#### Mold (rack)

People were very impressed with the recent improvements.  
Researching and developing a new type of rack had been great idea.

People were very impressed with the recent improvements.  
Researching and developing a new type of mold had been great idea.

The pottery and ceramics factory had been doing very well lately.  
Researching and developing a new type of rack had been great idea.

The pottery and ceramics factory had been doing very well lately.  
Researching and developing a new type of mold had been great idea.

#### Nails (bolts)

It seemed that they worked well as a team, but not so well individually.  
He was worried but she assured him that bolts like that would be fine.

It seemed that they worked well as a team, but not so well individually.  
He was worried but she assured him that nails like that would be fine.

The apprentice carpenter often looked to his mentor for guidance.  
He was worried but she assured him that bolts like that would be fine.

The apprentice carpenter often looked to his mentor for guidance.  
He was worried but she assured him that nails like that would be fine.

#### Panel (crowd)

Apparently she was only allowed to be there for fifteen minutes.  
She could hardly see the crowd because of the direction of the lighting.

Apparently she was only allowed to be there for fifteen minutes.  
She could hardly see the panel because of the direction of the lighting.

This was her last chance to impress them on the talent show.  
She could hardly see the crowd because of the direction of the lighting.

This was her last chance to impress them on the talent show.  
She could hardly see the panel because of the direction of the lighting.

#### Pupil (mouth)

The man had been investigating for hours and was about to give up.  
He couldn't figure out why the mouth was acting so strangely.

The man had been investigating for hours and was about to give up.  
He couldn't figure out why the pupil was acting so strangely.

The specialist couldn't give the patient a sensible medical explanation.  
He couldn't figure out why the mouth was acting so strangely.

The specialist couldn't give the patient a sensible medical explanation.  
He couldn't figure out why the pupil was acting so strangely.

#### Ruler (graph)

They were all pretty disappointed with how things had worked out.  
The hope was that this graph would better than the previous one.

They were all pretty disappointed with how things had worked out.  
The hope was that this ruler would better than the previous one.

These cabinet meetings were always tedious and full of deception.  
The hope was that this graph would better than the previous one.

These cabinet meetings were always tedious and full of deception.  
The hope was that this ruler would better than the previous one.

#### Sense (light)

The women had been working on the project for weeks.  
They knew that combining the two kinds of light would give them the best results.

The women had been working on the project for weeks.  
They knew that combining the two kinds of sense would give them the best results.

The physicists made a good team, one was intelligent and one was practical.  
They knew that combining the two kinds of light would give them the best results.

The physicists made a good team, one was intelligent and one was practical.  
They knew that combining the two kinds of sense would give them the best results.



Yard (gate)

The man was looking at the houses in the older part of town.  
He told his son that there was a gate between each one.

The man was looking at the houses in the older part of town.  
He told his son that there was a yard between each one.

The farm had closely packed milking stalls which made his job much quicker.  
He told his son that there was a gate between each one.

The farm had closely packed milking stalls which made his job much quicker.  
He told his son that there was a yard between each one.

## Appendix 3: Materials for Semantic Prediction Experiment

Below are the materials used in the Semantic Prediction experiment. The items are presented here in the following order:

Context 1 (a) Predicted  
Context 1 (an) Unpredicted  
Context 2 (an) Predicted  
Context 2 (a) Unpredicted

For the snowman's eyes the children used two pieces of coal, and for its nose they used a carrot from the fridge.

For the snowman's eyes the children used two pieces of coal, and for its nose they used an apple from the fridge.

The old wives tale says that if you want to keep the doctor away then you should eat an apple a day.

The old wives tale says that if you want to keep the doctor away then you should eat a carrot a day.

Katie did not like to say words with the letter s because she spoke with a lisp and was embarrassed.

Katie did not like to say words with the letter s because she spoke with an accent and was embarrassed.

It was difficult to understand the foreign professor because he had an accent when he spoke.

It was difficult to understand the foreign professor because he had a lisp when he spoke.

The bakery did not accept credit cards so Peter would have to write a cheque to the owner.

The bakery did not accept credit cards so Peter would have to write an apology to the owner.

Jack was very sorry for what he had said to Sarah and he knew that he owed her an apology because she was still upset.

Jack was very sorry for what he had said to Sarah and he knew that he owed her a cheque because she was still upset.

The day was breezy so the boys went outside to fly a kite in the park.

The day was breezy so the boys went outside to fly an aeroplane in the park.

Although the idea of flight was as old as the hills, the Wright Brothers were the first people to build an aeroplane that actually flew.

Although the idea of flight was as old as the hills, the Wright Brothers were the first people to build a kite that actually flew.

The tweeting in the treetops sounded like a bird to Melissa.  
The tweeting in the treetops sounded like an echo to Melissa.  
Surrounded by mountains, Jennifer shouted across the valley and heard an echo in the distance.  
Surrounded by mountains, Jennifer shouted across the valley and heard a bird in the distance.

Laura hated baiting the hook, but she knew that it was the only way to catch a fish without using a lure.  
Laura hated baiting the hook, but she knew that it was the only way to catch an octopus without using a lure.  
When the scuba diver saw the tentacle, he quickly realised that the creature under the rock was an octopus in hiding.  
When the scuba diver saw the tentacle, he quickly realised that the creature under the rock was a fish in hiding.

Samantha had no desire to enter the working world, so decided to stay at the university and remain a student for a few more semesters.  
Samantha had no desire to enter the working world, so decided to stay at the university and remain an artist for a few more semesters.  
On a street corner in Paris, Alice had her portrait painted by an artist for three Euros.  
On a street corner in Paris, Alice had her portrait painted by a student for three Euros.

James is an avid reader so, for his birthday, his sister decided to give him a book about Africa.  
James is an avid reader so, for his birthday, his sister decided to give him an education about Africa.  
Her grandfather always stressed how important it was for Sophie to go to school, because he had never had the opportunity to receive an education while growing up.  
Her grandfather always stressed how important it was for Sophie to go to school, because he had never had the opportunity to receive a book while growing up.

Without the floats and the marching bands it wasn't much of a parade this year.  
Without the floats and the marching bands it wasn't much of an election this year.  
When the representative retired in the middle of his term, the state was forced to hold an election in his district.  
When the representative retired in the middle of his term, the state was forced to hold a parade in his district.

Whenever Josh had too much to drink, he became belligerent and would try to start a fight with someone in the bar.  
Whenever Josh had too much to drink, he became belligerent and would try to start an affair with someone in the bar.  
When her husband started staying late at the office every night, Kirsty began to suspect that he was having an affair with someone at work.  
When her husband started staying late at the office every night, Kirsty began to suspect that he was having a fight with someone at work.

Andrew complained that the only place he could afford was the 99-cent store, after his grandmother gave him a dollar for helping her.

Andrew complained that the only place he could afford was the 99-cent store, after his grandmother gave him an allowance for helping her.

Instead of giving her money whenever she asked for it, Rachel's father thought that it would be a real lesson in economics for her if every week she received an allowance to spend as she wished.

Instead of giving her money whenever she asked for it, Rachel's father thought that it would be a real lesson in economics for her if every week she received a dollar to spend as she wished.

Harry saw smoke billowing out of his neighbours' home and immediately got on the phone to report a fire at their address.

Harry saw smoke billowing out of his neighbours' home and immediately got on the phone to report an emergency at their address.

When she had filled out the paperwork, Jessica had to write down the name of someone who could be contacted in case of an emergency while she was at work.

When she had filled out the paperwork, Jessica had to write down the name of someone who could be contacted in case of a fire while she was at work.

I wanted to put my coat in the wardrobe but I could not find a hanger anywhere in the house.

I wanted to put my coat in the wardrobe but I could not find an iron anywhere in the house. Charlie's wife told him that she was tired of pressing his shirts and that he would have to learn how to use an iron one of these days.

Charlie's wife told him that she was tired of pressing his shirts and that he would have to learn how to use a hanger one of these days.

When we saw the finish line and the cones sectioning off part of the road, we knew that there must have been a race earlier that day.

When we saw the finish line and the cones sectioning off part of the road, we knew that there must have been an accident earlier that day.

No matter how safely you drive your car, chances are someday you'll be involved in an accident with another vehicle.

No matter how safely you drive your car, chances are someday you'll be involved in a race with another vehicle.

Instead of recording a live album, the band decided they would have more control if they recorded in a studio without a lot of background noise.

Instead of recording a live album, the band decided they would have more control if they recorded in an apartment without a lot of background noise.

Mary had finally decided to buy a house in the suburbs, after a year of renting an apartment in the city.

Mary had finally decided to buy a house in the suburbs, after a year of renting a studio in the city.

When I called his house John was not home, but his mother said she could relay a message to him.

When I called his house John was not home, but his mother said she could relay an invitation to him.

Sue had wanted to go to Tim's birthday party but she was still waiting for an invitation from him.

Sue had wanted to go to Tim's birthday party but she was still waiting for a message from him.

You never forget how to ride a bicycle once you've learned.

You never forget how to ride an elephant once you've learned.

The highlight of Paul's trip to India was when he got to ride an elephant in the parade.

The highlight of Paul's trip to India was when he got to ride a bicycle in the parade.

As he walked past the corner shop on his way to work, Ian skimmed the headlines and decided to go ahead and buy a newspaper when he saw what the forecast was.

As he walked past the corner shop on his way to work, Ian skimmed the headlines and decided to go ahead and buy an umbrella when he saw what the forecast was.

Because it frequently rains in London, it's a good idea to always carry an umbrella with you.

Because it frequently rains in London, it's a good idea to always carry a newspaper with you.

When the pipe broke in the bathroom, Anne looked through the phonebook to find a plumber who could come take a look at the job.

When the pipe broke in the bathroom, Anne looked through the phonebook to find an architect who could come take a look at the job.

Frank wanted to design a very modern house so he sought advice from an architect to select the fixtures.

Frank wanted to design a very modern house so he sought advice from a plumber to select the fixtures.

The Hell's Angel rolled up the sleeve of his leather jacket to show the girls where he had got a tattoo the previous night.

The Hell's Angel rolled up the sleeve of his leather jacket to show the girls where he had got an infection the previous night.

Because Derek did not clean his wound properly, he ended up getting an infection on his leg.

Because Derek did not clean his wound properly, he ended up getting a tattoo on his leg.

Older children often have a harder time than younger children dealing with the loss of a parent who has cared for them.

Older children often have a harder time than younger children dealing with the loss of an adult who has cared for them.

In order to get into a 12A film, children under 12 must be accompanied by an adult or legal guardian.

In order to get into a 12A film, children under 12 must be accompanied by a parent or legal guardian.

Jeffrey posted the letter without a stamp so the post office would not deliver it.  
Jeffrey posted the letter without an envelope so the post office would not deliver it.  
When Wendy went to pay for the birthday card, the cashier pointed out that she had forgotten to pick up an envelope along with it.  
When Wendy went to pay for the birthday card, the cashier pointed out that she had forgotten to pick up a stamp along with it.

There are always ways to improve your game, whether you're an advanced tennis player or merely a beginner starting out.

There are always ways to improve your game, whether you're an advanced tennis player or merely an amateur starting out.

The professional photographer was disqualified from the photo contest because the rules explicitly state that you must be an amateur in order to compete.

The professional photographer was disqualified from the photo contest because the rules explicitly state that you must be a beginner in order to compete.

Before little Gloria went to sleep, she wanted to hear a story from her mother.

Before little Gloria went to sleep, she wanted to hear an excuse from her mother.

Amelia did not want to go to the bar with her friends after seeing the film, so she made up an excuse for why she needed to go home early.

Amelia did not want to go to the bar with her friends after seeing the film, so she made up a story for why she needed to go home early.

Marie wanted to sample a tiny bit of the sauce so she daintily dipped a finger into the pot.

Marie wanted to sample a tiny bit of the sauce so she daintily dipped an onion into the pot.

At first Victoria did not know why her brother was crying over the sink, but then she noticed that he had just sliced an onion with his knife.

At first Victoria did not know why her brother was crying over the sink, but then she noticed that he had just sliced a finger with his knife.

Jane's first client was a failure, but her second was a success and made her a lot of money.

Jane's first client was a failure, but her second was an author and made her a lot of money.

After Joanne's first book was published, she finally felt like she could call herself an author in front of her peers.

After Joanne's first book was published, she finally felt like she could call herself a success in front of her peers.

From a young age Steven had an interest in filmmaking and had always dreamed of being a director when he grew up.

From a young age Steven had an interest in filmmaking and had always dreamed of being an actor when he grew up.

David had moved to Hollywood in hopes of becoming an actor but so far he had only done a few commercials.

David had moved to Hollywood in hopes of becoming a director but so far he had only done a few commercials.

Claire's father had always joked that two is company but three is a crowd for him.  
Claire's father had always joked that two is company but three is an audience for him.  
Carole enjoyed singing in private, but was scared to death at the thought of performing in front of an audience at the concert.  
Carole enjoyed singing in private, but was scared to death at the thought of performing in front of a crowd at the concert.

The doctors would not allow Caroline to walk so soon after her surgery, so the only way for her to move from place to place was in a wheelchair accompanied by a nurse.  
The doctors would not allow Caroline to walk so soon after her surgery, so the only way for her to move from place to place was in an elevator accompanied by a nurse.  
Dominic preferred climbing stairs over riding in an elevator because it gave him a chance get a little exercise.  
Dominic preferred climbing stairs over riding in a wheelchair because it gave him a chance get a little exercise.

Hannah wanted to live in a small town, but her husband preferred to live closer to a city because of his job.  
Hannah wanted to live in a small town, but her husband preferred to live closer to an airport because of his job.  
The pilot had to make an emergency landing in the middle of the desert because he was nowhere near an airport or a safe place to land.  
The pilot had to make an emergency landing in the middle of the desert because he was nowhere near a city or a safe place to land.

Ever since Mr. Barnes had moved away, Bobby's football team had been left without a coach and thus lost all their games.  
Ever since Mr. Barnes had moved away, Bobby's football team had been left without an offense and thus lost all their games.  
Although the basketball team's defence was very strong, they did not have much of an offense since the middle of the season.  
Although the basketball team's defence was very strong, they did not have much of a coach since the middle of the season.

The guys didn't know what to call their band, so Tom told everyone to come up with a name and they would vote.  
The guys didn't know what to call their band, so Tom told everyone to come up with an idea and they would vote.  
The group had been brainstorming all day but they still didn't have an idea for their project.  
The group had been brainstorming all day but they still didn't have a name for their project.

Sandra decided that she would work for a cruise line for a year before starting college, although she had never been on a ship and did not know how to swim.  
Sandra decided that she would work for a cruise line for a year before starting college, although she had never been on an island and did not know how to swim.  
Being from the mainland, Karen never got used to the feeling of living on an island and being so removed from everything.  
Being from the mainland, Karen never got used to the feeling of living on a ship and being so removed from everything.

Every time they went for walks, Sylvia's dog Rex would break into a run as soon as he spotted a cat, but luckily Rex couldn't climb trees.

Every time they went for walks, Sylvia's dog Rex would break into a run as soon as he spotted an owl, but luckily Rex couldn't climb trees.

As Elizabeth climbed the ladder in the barn, she heard a hooting sound coming from the rafters and looked up to see an owl staring down at her.

As Elizabeth climbed the ladder in the barn, she heard a hooting sound coming from the rafters and looked up to see a cat staring down at her.

As the men stumbled across the desert, they thought they saw an oasis on the horizon, but it turned out to be a mirage so they continued walking.

As the men stumbled across the desert, they thought they saw an oasis on the horizon, but it turned out to be an ocean so they continued walking.

The first time Timmy saw the Pacific, he thought it was a lake, but his brother laughed and explained to him that it was an ocean and he should be careful of the waves.

The first time Timmy saw the Pacific, he thought it was a lake, but his brother laughed and explained to him that it was a mirage and he should be careful of the waves.

Georgina did not deal well with praise and had never really learned how to take a compliment from her superiors.

Georgina did not deal well with praise and had never really learned how to take an order from her superiors.

The waitress at the next table was trying to scribble everything down, but it was obvious that she did not know how to take an order from her customers.

The waitress at the next table was trying to scribble everything down, but it was obvious that she did not know how to take a compliment from her customers.

While Natasha was strolling through the colourful gardens, she reached down and picked a flower up off the ground.

While Natasha was strolling through the colourful gardens, she reached down and picked an animal up off the ground.

Isabel wasn't sure if the Venus flytrap was classified as a plant or as an animal but it seemed to have characteristics of both.

Isabel wasn't sure if the Venus flytrap was classified as a plant or as a flower, but it seemed to have characteristics of both.

Mike thought that glass was a solid but when he held it under the Bunsen burner it melted and turned into a liquid that could be sculpted.

Mike thought that glass was a solid but when he held it under the Bunsen burner it melted and turned into an element that could be sculpted.

The chemistry teacher explained to the class that water and salt were both compounds but that nitrogen was just an element that hadn't been combined.

The chemistry teacher explained to the class that water and salt were both compounds but that nitrogen was just a liquid that hadn't been combined.



Because they were playing baseball so close to the house, the children ended up shattering a window and destroying a nest that some birds had built on the window sill.

Because they were playing baseball so close to the house, the children ended up shattering an egg and destroying a nest that some birds had built on the window sill.

Violet was just learning how to make an omelette, so her father began by showing her how to crack an egg in the kitchen.

Violet was just learning how to make an omelette, so her father began by showing her how to crack a window in the kitchen.

With his khaki shorts, his loud Hawaiian shirt, and the camera around his neck, you could definitely tell that the man was a tourist in the country.

With his khaki shorts, his loud Hawaiian shirt, and the camera around his neck, you could definitely tell that the man was an alien in the country.

Bob claims to be from outer space, but nobody believes that he is actually an alien from another planet.

Bob claims to be from outer space, but nobody believes that he is actually a tourist from another planet.